

2012

Improving Iris Recognition through Quality and Interoperability Metrics

Nathan D. Kalka
West Virginia University

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Kalka, Nathan D., "Improving Iris Recognition through Quality and Interoperability Metrics" (2012).
Graduate Theses, Dissertations, and Problem Reports. 4872.
<https://researchrepository.wvu.edu/etd/4872>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

Improving Iris Recognition through Quality and Interoperability Metrics

Nathan D. Kalka

Dissertation submitted to the
Benjamin M. Statler College of Engineering and Mineral Resources
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in
Computer and Information Science

Bojan Cukic, Ph.D., Chair
Arun Ross, Ph.D.
Lawrence Hornak, Ph.D.
Keith Morris, Ph.D.
Donald Adjero, Ph.D.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia
2012

Keywords: Iris Recognition, Iris Segmentation, Image Quality, Segmentation Evaluation,
Segmentation Rectification, Iris Image Interoperability

Copyright ©2012 Nathan D. Kalka

Abstract

Improving Iris Recognition through Quality and Interoperability Metrics

Nathan D. Kalka

The ability to identify individuals based on their iris is known as iris recognition. Over the past decade iris recognition has garnered much attention because of its strong performance in comparison with other mainstream biometrics such as fingerprint and face recognition. Performance of iris recognition systems is driven by application scenario requirements. Standoff distance, subject cooperation, underlying optics, and illumination are a few examples of these requirements which dictate the nature of images an iris recognition system has to process. Traditional iris recognition systems, dubbed “stop and stare”, operate under highly constrained conditions. This ensures that the captured image is of sufficient quality so that the success of subsequent processing stages, segmentation, encoding, and matching are not compromised. When acquisition constraints are relaxed, such as for surveillance or iris on the move, the fidelity of subsequent processing steps lessens.

In this dissertation we propose a multi-faceted framework for mitigating the difficulties associated with non-ideal iris. We develop and investigate a comprehensive iris image quality metric that is predictive of iris matching performance. The metric is composed of photometric measures such as defocus, motion blur, and illumination, but also contains domain specific measures such as occlusion, and gaze angle. These measures are then combined through a fusion rule based on Dempster-Shafer theory. Related to iris segmentation, which is arguably one of the most important tasks in iris recognition, we develop metrics which are used to evaluate the precision of the pupil and iris boundaries. Furthermore, we illustrate three methods which take advantage of the proposed segmentation metrics for rectifying incorrect segmentation boundaries. Finally, we look at the issue of iris image interoperability and demonstrate that techniques from the field of hardware fingerprinting can be utilized to improve iris matching performance when images captured from distinct sensors are involved.

This work is dedicated to my family for their encouragement and support

*To Alicia, for her unequivocal support, patience, and understanding throughout the duration
of this work*

Acknowledgments

First, and foremost, I would like to offer my deepest thanks to my advisor and committee Chair Dr. Bojan Cukic. He has provided me with many unique learning opportunities throughout the duration of this work. The guidance, wisdom, and knowledge he has provided me has been invaluable. I am truly indebted to him.

To my graduate committee members: Dr. Larry Hornak, Dr. Arun Ross, Dr. Keith Morris and Dr. Donald Adjero. The commentary, direction, and discussion they have provided me has been invaluable for strengthening this work.

I would also like to acknowledge my friends and colleagues: Dr. Jeremy Dawson, Dr. Mahos Bourlai, Dr. Simona Crihalmeanu and Dr. Nick Bartlow for invaluable discussions and ideas related to this work. Without their support, this work would not have been possible.

Contents

1	Introduction	3
1.1	Motivation	3
1.2	Goal	4
1.3	Contributions	6
1.4	Organization	7
2	Background	9
2.1	Iris Anatomy	9
2.2	Iris Recognition	11
2.2.1	Daugman’s Framework	12
2.2.2	Wildes’ Framework	16
2.3	Image Acquisition	19
2.3.1	Commercial Acquisition Systems	19
2.3.2	Unconstrained Subject Presentation	22
2.3.3	Iris Image Quality	24
2.4	Segmentation	27
2.4.1	Orthogonal View Processing	27
2.4.2	Non-Orthogonal View	29
2.4.3	Segmentation Evaluation	32
3	Iris Image quality	35
3.1	Introduction	35
3.2	Non-ideal data set collection and Experimentation	37
3.2.1	Defocus Blur	38
3.2.2	Off-Angle	39
3.2.3	Occlusion	40
3.2.4	Lighting Variation	41
3.2.5	Iris Resolution	43
3.2.6	Discussion	45

3.3	Estimation of Individual Quality Factors	45
3.3.1	Defocus	46
3.3.2	Motion	47
3.3.3	Off-Angle	50
3.3.4	Occlusion	51
3.3.5	Specular Reflection	52
3.3.6	Lighting Variation	54
3.3.7	Pixel Counts	54
3.4	Combining Quality Factors	55
3.4.1	Quality Score Normalization	57
3.4.2	Dempster Shafer Theory Applied to Quality Assessment	58
3.4.3	Evidence Fusion Examples Based on Murphy's Rule	59
3.5	Numerical Results	61
3.5.1	Quality Factor Distributions	61
3.5.2	CASIA 3.0, WVU, and ICE 1.0 Performance	67
3.6	Performance Comparison	68
3.6.1	Experimental Results	69
3.7	Chapter Summary	71
4	Segmentation Evaluation	73
4.1	Introduction	73
4.2	How do we evaluate segmentation?	75
4.3	Approach	76
4.3.1	Geometric Measures	77
4.3.2	Statistical Intensity Measure	79
4.3.3	Edge Measure	84
4.4	Experimental Design	86
4.5	Experimental Results	90
4.5.1	Masek Segmentation	91
4.5.2	IDO Segmentation	95
4.5.3	Zuo et al. Segmentation	99
4.5.4	Evaluating the Overall Segmentation Result	103
4.5.5	Filtering Iris Match Scores Utilizing Segmentation Evaluation	106
4.6	Chapter Summary	109
5	Segmentation Rectification	111
5.1	Introduction	111
5.2	Segmentation Rectification Strategy - S1	114

5.2.1	Experimental Results	114
5.3	Segmentation Rectification Strategy - S2	117
5.3.1	Experimental Results	119
5.4	Segmentation Rectification Strategy - S3	121
5.4.1	Experimental Results	125
5.5	Chapter Summary	128
6	Iris Image Interoperability	129
6.1	Introduction	129
6.2	Approach	131
6.2.1	Sensor Identification	132
6.2.2	Wavelet Based Denoising Algorithm	133
6.3	Experimental Design	134
6.3.1	Chimerical Data Set Construction	134
6.3.2	Interoperability - 2009 Biometric Collection Project Data Set	136
6.4	Results - Hardware Fingerprinting	139
6.4.1	Unit Level Sensor Identification	139
6.4.2	Unit and Brand Level Sensor Identification	140
6.4.3	Impact of JPEG Compression	142
6.4.4	Interoperability - 2009 Biometric Collection Project Data Set	144
6.5	Chapter Summary	146
7	Conclusions	147
7.1	Conclusions	147
7.2	Future Work	149

List of Figures

1.1	Framework for mitigating some of the challenges associated with non-ideal iris. The shaded dark blue regions of the framework represent what we are proposing in this dissertation. Specifically, during acquisition, quality assessment (1.1) can be utilized to cull images that would otherwise be detrimental to subsequent processing stages of the iris recognition system. Following acquisition, segmentation is attempted on the captured image which isolates iris texture from surrounding periocular image structures. Here, we are proposing the addition of a segmentation evaluation (2.1) and rectification block (2.2 - 2.2b). These blocks will determine whether the segmentation process has succeeded or failed. In the latter case, the rectification block will attempt to correct for failed iris localization. Finally, when the issue of interoperability arises, block 4.1 is utilized to maintain match rates when processing images captured by different iris sensors.	5
2.1	Illustration of human iris imaged under near-infrared (700-900nm). Note the surrounding periocular structures such as eyelids, eyelashes, and the sclera. Eyelids and eyelashes can have a negative impact on the iris segmentation process when excessively occluding the iris region.	10
2.2	Illustration of the mapping between cartesian coordinates and polar coordinates. $(x_p(\theta), y_p(\theta))$ and $(x_s(\theta), y_s(\theta))$ represent a pair of points on the pupil and iris boundaries at angle θ and radius r respectively. This process effectively makes subsequent recognition steps robust to geometric distortions and linear deformations associated with pupil dilation and constriction.	15

2.3	Illustration of circular Hough transform for finding the pupillary boundary. The input image (left column) is first smoothed, followed by thresholding of the gradient magnitude, in order to generate an edge map (center column). Finally, equations 2.8-2.10 are used to generate a Hough accumulator matrix (right column - Accumulator matrix generated with $r = 42$). In implementation, the search iterates over a range of radii and an accumulator matrix is generated for each radius.	17
2.4	Illustration of commercial iris acquisition systems. Top row: LG Iris Access 4000, Iris Guard AD100, Oki Irispass-M. Bottom row: Sarnoff Passport, Aoptix Insight, and GRI's HBOX-V. Note that systems on the bottom row impose fewer constraints on the user.	20
3.1	(a) Effect of defocus blur on Log-Gabor encoding technique. Distance 1 corresponds to the ideal images and 2, 3, and 4 represent when the subject is $\frac{1}{8}$, $\frac{1}{4}$, and $\frac{3}{4}$ inches from the edge of the chin rest respectively. (b) & (c) are sample images taken from $\frac{1}{4}$ and $\frac{3}{4}$ inches from the edge of the chin rest. . .	39
3.2	(a) Performance of Log-Gabor filter-based recognition algorithm vs. increasing yaw angle. (b) and (c) are sample images from the data set.	40
3.3	(a) Degradation of recognition performance due to varying levels of occlusion. (b) and (c) represent occluded images from a subject for the level 3 and level 4 groups respectively.	41
3.4	Example of images with non-uniform lighting conditions from the WVU data set.	42
3.5	(a) Influence of lighting deficiency on Log-Gabor based encoding.(b) and (c) represent sample images from a subject when the light source is 3 and 4 feet behind the camera respectively.	43
3.6	(a) Dependence of recognition performance on iris diameter. (b), (c), (d) are images from a subject with iris diameter of 223, 168, and 120 pixels respectively.	44
3.7	Sample images from WVU data set and their band-pass filter responses. The local assessment in (a) is performed based on the pupil and iris localization. The total width of the assessment region is $\frac{1}{3}$ of the iris region. The assessment region begins from $\frac{1}{7}$ of the distance from the pupil to iris boundary.	47
3.8	Motion estimation block diagram	49
3.9	Log magnitude slice before and after b-spline smoothing.	49
3.10	Off-Angle Estimation (a) An iris image that is out of plane by 30° which is assigned during data collection. (b) The image in (a) brought back in to frontal view by projective transformation. (c) A plot of the integro-differential operator over various pitch and yaw angles.	51

3.11	Occlusion mask generation.	53
3.12	Illustration of specular reflection mask generation.	53
3.13	Sample Images from WVU, ICE 1.0, and CASIA 3.0 data sets	60
3.14	We classify images as incorrect if pupil and iris segmentation boundaries do not fall completely along their respective boundaries. (a), (b), (c) represent correct localization results from CASIA 3.0, WVU, and ICE 1.0 respectively. (d), (e), (f) represent incorrect localization from CASIA 3.0, WVU, and ICE 1.0 respectively.	62
3.15	Defocus and motion frequencies for CASIA 3.0, WVU, and ICE	63
3.16	Occlusion and lighting frequencies for CASIA 3.0, WVU, and ICE	64
3.17	Specular and pixel count frequencies for CASIA 3.0, WVU, and ICE	65
3.18	Overall quality scatter plots for CASIA 3.0, WVU, ICE 1.0 data sets	66
3.19	Verification performance for all three data sets.	68
3.20	CASIA V3.0 Performance prediction comparision	69
3.21	WVU Performance prediction comparison	70
4.1	Four types of segmentation results (a) Correctly segmented image (b) Failed pupil segmentation (c) Failed iris segmentation (d) Failed pupil and iris segmentation.	74
4.2	Illustration of annular distance ratio. The image in the top row is grossly inaccurate with respect to the iris boundary. The corresponding distance between the boundary centers is over 40 pixels while the boundary centers in the bottom row image are offset by only 13 pixels.	79
4.3	(1.) An iris image and its overlaid segmentation boundaries. (2.) Mask out all regions outside of the segmentation. (3.) Compute the intensity histogram of the unmasked region and estimate parameters for each model. For this specific image, $Bin(P_t) = 24$, $Bin(I_t) = 123$, $\hat{\Theta} = 4.75$, $\hat{\mu} = 73.93$, and $\hat{\sigma} = 19.39$. (4.) Every pixel within the estimated pupil segmentation is classified as belonging to the pupil or not. Red pixels were classified as non-pupil pixels; blue circle corresponds to the estimated pupil segmentation boundary from Masek's algorithm.	81
4.4	Pupil under-segmentation estimation (a) Under segmented pupil (b) Estimated under-segmentation.	84

4.5	Illustration of the region edge measure. The left column represents a correct segmentation for both pupil and iris boundaries while the image in the right column has an inferior iris boundary. The red circles represent points on the estimated pupil/iris boundary at specific sampling angles. The iris boundary sampling angles are 10,15,20 for the left hemisphere and 170,165,160 for the right hemisphere in both images. The green circles correspond to the location of the edge region detected by equation 4.8. Note that the tables below the images represent the average edge scores for the sampling angles on the left and right hemisphere of the boundary respectively.	86
4.6	Sample imagery from each data set: (a) ICE, (b) WVU, and (c) QFIRE. Images discarded during frame extraction of the QFIRE videos are illustrated in (d). It is important to note that all three data sets are not indicative of operational data because quality control was not employed during collection. Thus, they are of particular interest for the focus of this work, on processing strategies for non-ideal iris.	87
4.7	Histograms of Tanimoto coefficient for both Masek and IDO algorithms. The Tanimoto coefficient, in this context, is utilized to characterize the ground truth for each segmentation algorithm. That is, ground truth for correct segmentation on average correspond to a coefficient of .95. On the other hand, segmentation below a coefficient of .95 correspond to failed segmentation. . .	89
4.8	Segmentation score distributions for Masek segmentation on WVU data (a) Average annular distance ratio distributions (b) Pupil-Over distributions (c) Pupil-under distributions (d) Average iris edge score distributions (e) Segmentation metric ROC.	92
4.9	Segmentation score distributions for Masek segmentation on ICE data (a) Average annular distance ratio distributions (b) Pupil-Over distributions (c) Pupil-under distributions (d) Average iris edge score distributions (e) Segmentation metric ROC.	94
4.10	ROC performance for the segmentation metrics generated from the Masek segmentation boundaries on the QFIRE data set.	96
4.11	Segmentation score distributions for IDO segmentation on ICE data (a) Average annular distance ratio distributions (b) Pupil-Over distributions (c) Pupil-under distributions (d) Average iris edge score distributions (e) Segmentation metric ROC.	98
4.12	Segmentation score distributions for IDO segmentation on ICE data (a) Average annular distance ratio distributions (b) Pupil-Over distributions (c) Pupil-under distributions (d) Average iris edge score distributions (e) Segmentation metric ROC.	100

4.13	ROC performance for the segmentation metrics generated from the IDO segmentation boundaries on the QFIRE data set.	101
4.14	Pupil / iris ROC curves for Zuo et. al segmentation (a) ICE, (b) WVU, and (c) QFIRE.	102
4.15	Illustration of the sequential C-SVM ensemble. Gross segmentation failures are detected with the model generated from annular distance ratios. If evaluated to be grossly inaccurate processing stops, otherwise the pupil features are evaluated within the next model. This process continues sequentially until the iris edge features are evaluated. If all models in the ensemble indicate successful segmentation, then the overall segmentation is classified as correct. Note that the dotted lines indicate optional processing steps since any model in the ensemble could detect an incorrect boundary, thus processing stops and subsequent models within the ensemble are not evaluated.	103
4.16	Iris recognition performance ROC's based on segmentation result evaluations. (a) Masek-WVU, (b) IDO-WVU (c) Masek-ICE, (d) IDO-ICE, (e) Masek-QFIRE, (f) IDO-QFIRE	108
5.1	ROC curves characterizing the performance of strategy S1 rectification performance across all data sets. Rectification strategy S1 is observed to increase iris matching performance for all data sets. However, there is less observed improvement for the QFIRE data resulting from overlap in segmentation failures between the baseline and rectification segmentation algorithms.	116
5.2	ROC curves characterizing the performance of strategy S2 rectification performance. Rectification strategy S2 is observed to increase iris matching performance for all data sets.	122
5.3	Illustration of pupil rectification process. (a) - Inferior segmentation overlayed on original image. (b) - Evaluation mask generated from the pupil over-segmentation and under-segmentation features. (c) - Edge map generated from original image. (d) - Edge map generated from the evaluation mask. . .	124
5.4	ROC curves characterizing the performance of strategy S3 rectification performance.	127
6.1	Illustration of reference templates generated from 5 training samples: (a) ICE-LG and (b) WVU-OKI.	133
6.2	Example match and non-match distributions with 32 training images for CASIAv3-OKI.	140
6.3	Unit Level Sensor identification as a function of training set size.	141

6.4	ICE1.0 example match and non-match distributions with 32 training images per sensor.	142
6.5	Cumulative match characteristic curve as a function of train/test scenarios when considering all seven sensors.	143
6.6	Rank-1 and Rank-2 Identification performance for all seven sensors as a function of JPEG quality level. In this illustration reference patterns were generated from 32 training images.	144
6.7	ROC curves characterizing the performance of single sensor and cross sensor performance. Note that in (b), hardware fingerprinting is used to ensure that matching is performed on imagery captured from the same sensor. As a result, single sensor performance is not degraded.	145

List of Tables

2.1	Survey of selected commercial iris recognition cameras. The majority of these sensors place constraints on the user for successful operation. The sensors in <i>italics</i> can operate at further distances or even while users are walking, resulting in a less intrusive experience.	21
3.1	Daugman's 8x8 convolution kernel	46
3.2	Estimated Factors for images in Fig. 3.13	60
3.3	Segmentation statistics for all three data sets.	61
3.4	CASIA 3.0, WVU, and ICE 1.0 Mean and Variance Quality Factor Scores . .	67
3.5	Quality interval ranges for each data set.	67
3.6	Summary of performance statistics using Log-Gabor encoding algorithm for the CASIA 3.0 data set and WVU data set.	68
3.7	Summary of performance statistics using Log-Gabor encoding algorithm for the ICE 1.0 data set.	68
4.1	Segmentation ground truth results for all algorithms across WVU, ICE, and QFIRE data sets. Note that the Zuo et al. algorithm failed to produce segmentation boundaries for 13 of the 583 in the QFIRE subset, which were excluded from subsequent analysis.	88
4.2	Segmentation metric distribution statistics for Masek segmentation on the QFIRE data set.	96
4.3	Segmentation metric distribution statistics for IDO segmentation on Q-FIRE.	101
4.4	Segmentation metric distribution statistics for Zuo's segmentation.	101
4.5	Annular Distance Ratios: C-SVM Ensemble Misclassifications. A G = actual good segmentation, A F = actual failed segmentation, P G = predicted good segmentation, P F = predicted failed segmentation	105

4.6	Annular Distance Ratios + Pupil Over/Under-Segmentation: C-SVM Ensemble Misclassifications. A G = actual good segmentation, A F = actual failed segmentation, P G = predicted good segmentation, P F = predicted failed segmentation	106
4.7	Annular Distance Ratio + Pupil Over/Under-Segmentation + Iris Edge: C-SVM Ensemble Misclassifications. A G = actual good segmentation, A F = actual failed segmentation, P G = predicted good segmentation, P F = predicted failed segmentation	107
5.1	Summary statistics for rectification strategy S1. The “Evaluated-Failed” column represents the total number of detected failed segmentations via the ensemble, while the “Failed Post-Rectification” column corresponds to iris images whose boundaries could not be rectified. Clearly, rectification strategy S1 improves iris matching performance as the observed EER and mean genuine hamming distance scores for all data sets decreases post rectification.	115
5.2	Summary statistics for rectification strategy S2. The “Pre-Rectification” column represents the total number of ground truth failed segmentations, while the “Post-Rectification” column corresponds to iris images whose boundaries could not be rectified through search operator conditioning.	120
5.3	Summary statistics for rectification strategy S3.	126
6.1	Train and test scenarios for all sensors.	136
6.2	Illustration of sample images and corresponding noise residuals from each database.	137
6.3	Train and test scenarios for the three commercial iris scanners.	138
6.4	Confusion matrix when training on 32 images per sensor.	139
6.5	Confusion matrix when training on 32 images per sensor with JPEG compression (Quality=75).	142
6.6	EER performance for single sensor and cross sensor experiments. Note that the number in parenthesis corresponds to the ERR with hardware fingerprinting is applied and probe data is unlabeled.	145

Chapter 1

Introduction

1.1 Motivation

The ability to identify or recognize individuals based on their iris is known as iris recognition [48]. Over the past decade iris recognition has garnered much attention, in part because of observed error rates [76], but also because of its exceptional performance over mainstream biometrics such as fingerprint and face recognition. Performance of iris recognition systems is driven in part by application scenario requirements. Standoff distance, subject cooperation, underlying optics, and illumination are just a few examples of these requirements which dictate the nature of images an iris recognition system will have to process [86]. Traditional iris recognition systems, dubbed “stop and stare”, impose acquisition constraints on the user such that the success of subsequent processing stages, namely, segmentation, encoding, and matching are not compromised. However, in the absence of such constraints or when constraints are intentionally lax, such as for surveillance [5, 6] or iris on the move [67], the fidelity of subsequent processing blocks may become questionable. Regardless of the

application, there is an underlying need for ensuring the integrity, or “hardening”, critical stages of an iris recognition system. For instance, quality assessment algorithms may be employed during acquisition to cull imagery that would otherwise cause subsequent processing stages to fail. Alternatively, one can envision quality assessment being utilized adaptively, ushering images to specific processing paths based on their characterization, presumably where they can be successfully dealt with. Related to segmentation, research has found that iris segmentation results are a dominant factor that drives iris recognition matching performance [29, 68]. As a result, the ability to automatically determine whether the segmentation block of an iris recognition system has succeeded or failed is of paramount importance when attempting to predict the outcome of matching. Whether a binary success / failure flag or a measure with higher granularity, currently existing algorithms do not explicitly evaluate iris segmentation. Therefore, without human inspection, the success of the segmentation block is largely unknown in most iris recognition systems.

1.2 Goal

In this dissertation we propose a multi-faceted framework, illustrated in Figure 1.1, for helping mitigate the aforementioned difficulties associated with non-ideal iris. Related to the acquisition process, we develop and investigate a comprehensive iris image quality metric that is predictive of iris matching performance. The metric is composed of photometric measures such as defocus, motion blur, and illumination, but is also composed of domain specific measures such as occlusion, and gaze angle. These measures are then combined through a fusion rule based on dempster-shafer theory. Related to iris segmentation, which is arguably

one of the most important tasks in iris recognition, we develop several segmentation metrics which are used to evaluate the fidelity of the pupil and iris boundaries. Furthermore, we illustrate three methods which take advantage of the proposed segmentation metrics which can be utilized for rectifying incorrect segmentation boundaries. Finally, we look at the issue of iris image interoperability and demonstrate that techniques from the field of hardware fingerprinting can be utilized to improve iris matching performance when images captured from multiple sensors are involved.

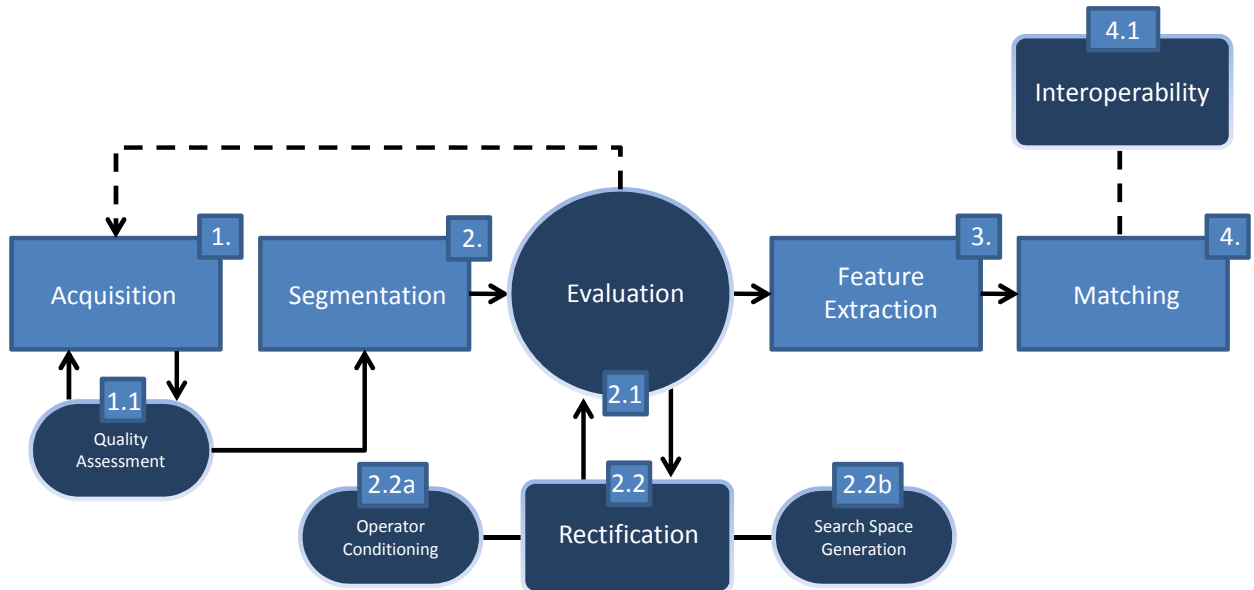


Figure 1.1: Framework for mitigating some of the challenges associated with non-ideal iris. The shaded dark blue regions of the framework represent what we are proposing in this dissertation. Specifically, during acquisition, quality assessment (1.1) can be utilized to cull images that would otherwise be detrimental to subsequent processing stages of the iris recognition system. Following acquisition, segmentation is attempted on the captured image which isolates iris texture from surrounding periocular image structures. Here, we are proposing the addition of a segmentation evaluation (2.1) and rectification block (2.2 - 2.2b). These blocks will determine whether the segmentation process has succeeded or failed. In the latter case, the rectification block will attempt to correct for failed iris localization. Finally, when the issue of interoperability arises, block 4.1 is utilized to maintain match rates when processing images captured by different iris sensors.

1.3 Contributions

This dissertation provides the following set of original contributions:

1. A comprehensive iris image quality metric that is predictive of iris matching performance. The metric is composed of photometric measures such as defocus, motion blur, and illumination, but is also composed of domain specific measures such as occlusion, and gaze angle. These measures are then combined through a fusion rule based on dempster-shafer theory.
2. Demonstration of the ability to reliably evaluate the success or failure of the iris segmentation routine. We analyze the estimated segmentation boundaries for both pupil and iris, and derive distance, shape, and edge metrics. These measures are then combined through a machine learning model which outputs a binary decision indicating the success or failure of estimated segmentation. These metrics are not tethered to any specific iris segmentation paradigm and as a result can be applied to many popular iris segmentation algorithms.
3. Demonstration of the ability to rectify incorrect segmentation boundaries when evaluated to be incorrect. We illustrate three methods which utilize the salient information generated by the segmentation evaluation methodology for the purpose of rectifying erroneous boundaries for both iris and pupil regions. The first method is a black box approach that utilizes the evaluation model to flag the use of a computationally more expensive segmentation routine to rectify inferior segmentation. The second method utilizes the evaluation measures by conditioning the search operator of the baseline segmentation algorithms. Finally, the third method is a pseudo-combination of the first

two methods in that we utilize the segmentation evaluation measures to condition and rectify inferior boundaries by augmenting a relatively weaker segmentation algorithm in order to make it stronger.

4. Demonstration of the application of hardware fingerprinting to iris sensors for increasing sensor interoperability. When images captured from multiple sensors are involved, we show that the application of hardware fingerprinting can be utilized to improve or at least maintain match rates.

1.4 Organization

The remainder of this dissertation is organized as follows. Chapter 2 provides a summary of related work regarding iris recognition, iris segmentation, iris image quality, segmentation evaluation, and iris image interoperability. Chapters 3-6 describe specific contributions in the areas of iris image quality, iris segmentation evaluation, iris segmentation rectification and iris image interoperability. Finally, Chapter 7 concludes the work by providing a summary of the accomplishments as well as future directions for research.

Chapter 2

Background

This chapter presents a succinct review of the related work in the fundamental areas of iris recognition. Relevant fields of interest with respect to this proposal include iris anatomy, image acquisition, iris image quality, and iris segmentation. The intent is to provide the reader with a basic understanding of the field.

2.1 Iris Anatomy

The iris is a thin, colored, radial structure within the eye that surrounds the pupil consisting mainly of smooth muscle. It is composed of an anterior layer known as the stroma and a posterior layer consisting of epithelial cells. The stroma connects to a circular group of muscle tissue called the sphincter which is responsible for contracting the pupil and a set of dilator muscles which are responsible for radially dilating the pupil. The iris and attached muscles are responsible for regulating the amount of light that reaches the retina.

The surface of the iris is dichotomized into a pupillary and ciliary region. These regions

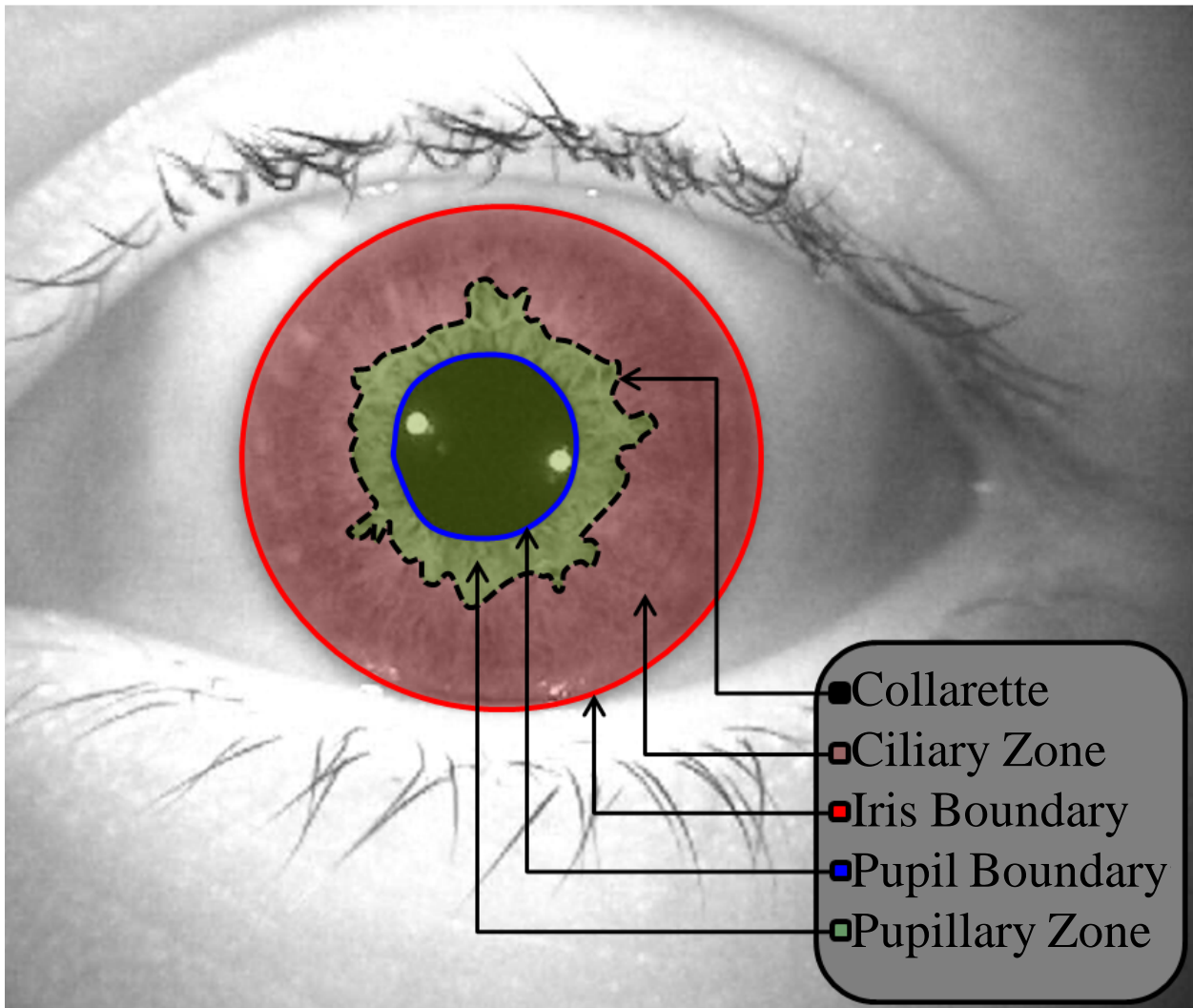


Figure 2.1: Illustration of human iris imaged under near-infrared (700-900nm). Note the surrounding periocular structures such as eyelids, eyelashes, and the sclera. Eyelids and eyelashes can have a negative impact on the iris segmentation process when excessively occluding the iris region.

are separated by the collarette, the thickest region in the iris, also characterized as the region where the dilator and sphincter muscles overlap. The pupillary region is the inner region which extends from the collarette to the edge region which forms the boundary of the darker

pupil. The ciliary region extends outward from the collarette to the outer edge of the iris which is surrounded by the sclera. An illustration of these regions can be found in fig. 2.1.

Literature on the iris and its rich structure suggests that it is determined randomly during prenatal morphogenesis of the eye. It is also believed that the iris is structurally diverse across the population [30, 78] and between genetically identical irides (left and right eye of the same individual) [30, 37].

2.2 Iris Recognition

Automated iris recognition was first proposed in a patent [37] by ophthalmologists, Flom and Safir. Although their system was not implemented, they illustrated concepts which are utilized in current iris recognition platforms. Their system captured images of the eye under adjustable monochromatic illumination at spatially separated positions. Light intensity was adjusted until the pupil reached a predetermined size for imaging, in effect, accounting for dilation and contraction. The position of the light source(s) varied such that elevation dependent features within the iris, such as radial furrows, could be imaged. Subsequent processing extracted features from the pupil and iris using image processing and pattern recognition methodologies. They describe the use of various forms of the Hough transform for descriptors of length, angular location, and direction of radial furrows, crypts, and pigment singularities. Additionally, the authors suggested descriptors based on correlation, high curvature, low order central moments, thresholding, and finite difference operators.

A working operational iris recognition system was presented by Daugman [32] in 1993 and patented [33] in 1994. His system encompasses four main processing blocks: (1) **Im-**

age acquisition pertains to imaging the periocular region to obtain a canonicalized image representation that is dominated by the iris; (2) *Segmentation* is the process of isolating iris tissue from surrounding image structures; (3) *Feature extraction* is the process of extracting unique identifiers from iris tissue, typically compact and invariant to distortions, that can be used for matching; (4) *Matching* compares extracted features for the purpose of identification or verification. Unlike Flom and Safir, Daugman provided quantitative analysis of his iris recognition system, focusing mainly on characterizing the statistical complexity of iris as a biometric. Detailed description of his framework as well as recent research on iris recognition is provided in the following sections.

2.2.1 Daugman's Framework

Daugman's initial work described a video acquisition system requiring subjects to position their eye within the camera field of view [32, 33]. Excessive eyelid occlusion was reduced by providing live feedback to cooperating subjects through an LCTV monitor or mirror. Images were acquired at a spatial resolution of 480x640 in the visible spectrum of light. An update to this system suggested using a monochrome CCD video camera with near-infrared illumination in the 700-900nm band [26, 30]. Near-infrared is invisible and unobtrusive to the human eye resulting in little impact on the size of the pupil (visible light can cause the pupil to dilate or constrict dependent upon light intensity). Furthermore, near-infrared illumination reveals rich detailed structure of irides that have strong pigmentation (such as dark brown eyes) which is not apparent under visible wavelengths since the melanin pigment in the iris absorbs most of the visible light while longer wavelengths are reflected [11, 27]. The high frequency power in the 2D Fourier spectrum is quantified for each frame, acting as

a rough focus assessment, which ensures that subsequent processing stages perform successfully.

After image acquisition, Daugman employs an efficient integro-differential operator for detecting the pupillary and iris boundaries [32]. It acts as a circular edge detector that iteratively searches through image space for a maximum contour integral derivative. Mathematically it is described as:

$$\max_{(r, x_0, y_0)} \left| G_\sigma(r) * \frac{\partial}{\partial r} \oint_{r, x_0, y_0} \frac{I(x, y)}{2\pi r} ds \right|, \quad (2.1)$$

where $G_\sigma(r)$ is a Gaussian smoothing function at scale σ , convolution is denoted by $*$, and (r, x_0, y_0) correspond to the radius and center coordinates which define a path for contour integration. Daugman constrains the angular search path, ds , to the left and right quadrants of the iris (vertical edge tuning) since heterogeneous image structures such as eyelids or eyelashes may occlude the upper and lower border of the iris. Subsequent processing searches the interior of the localized iris for the pupil limbus using equation 2.1, while constraining the angular search path to the upper 270° to avoid specular reflection induced by the light source. After both boundaries have been estimated, equation 2.1 is modified from circular to arcuate, for subsequent detection of the upper and lower eyelids which are modeled as splines [26].

After isolating the iris from adjacent periocular image structures, it is normalized to a doubly dimensionless polar coordinate system. Each cartesian point, (x, y) , on the iris is assigned a pair of real dimensionless coordinates (r, θ) , where $r \in [0, 1]$ is the radius variable and $\theta \in [0, 2\pi]$ is the angular variable. Mathematically this representation from cartesian to

polar is expressed as:

$$I(x(r, \theta), y(r, \theta)) \longrightarrow I(r, \theta), \quad (2.2)$$

$$x(r, \theta) = (1 - r)x_p(\theta) + rx_s(\theta), \quad (2.3)$$

$$y(r, \theta) = (1 - r)y_p(\theta) + ry_s(\theta), \quad (2.4)$$

where $x(r, \theta)$ and $y(r, \theta)$ represent linear combinations of boundary points around the pupil and iris boundary. Fig. 2.2 provides a detailed illustration of this process. Normalization ensures that feature extraction is robust to geometric distortions and linear deformations associated with pupil dilation and constriction. (Proponents of alternative normalization models argue that this representation is only suitable as an approximation. Pupillary functions such as dilation and constriction cause elastic deformations to surrounding iris tissue which may be represented better through nonlinear normalization models [54, 109, 116, 118].)

Daugman employs 2-dimensional (2D) Gabor filters for extracting features from isolated iris texture. The filters are convolved with the normalized iris data representation to extract image texture information. The coefficients associated with the phase response from each filter is quantized into a pair for bits depending on the *sgn* of the 2D integral in equation 2.5:

$$h_{\{Re, Im\}} = \underset{\{Re, Im\}}{\text{sgn}} \int_{\rho} \int_{\phi} I(\rho, \phi) e^{-i\omega(\theta_0 - \phi)} \cdot e^{\frac{-(r_0 - \rho)^2}{\alpha^2}} e^{\frac{-(\theta_0 - \phi)^2}{\beta^2}} \rho d\rho d\phi, \quad (2.5)$$

where $I(\rho, \phi)$ is the normalized raw iris representation; α , β , and ω represent the size and frequency parameters of the wavelet; (r_0, θ_0) define the polar coordinates. This process is repeated across the entire iris region with different wavelet scales, frequencies, and orientations to extract a 2048 bit compact binary representation, called an iris code. The resulting binary

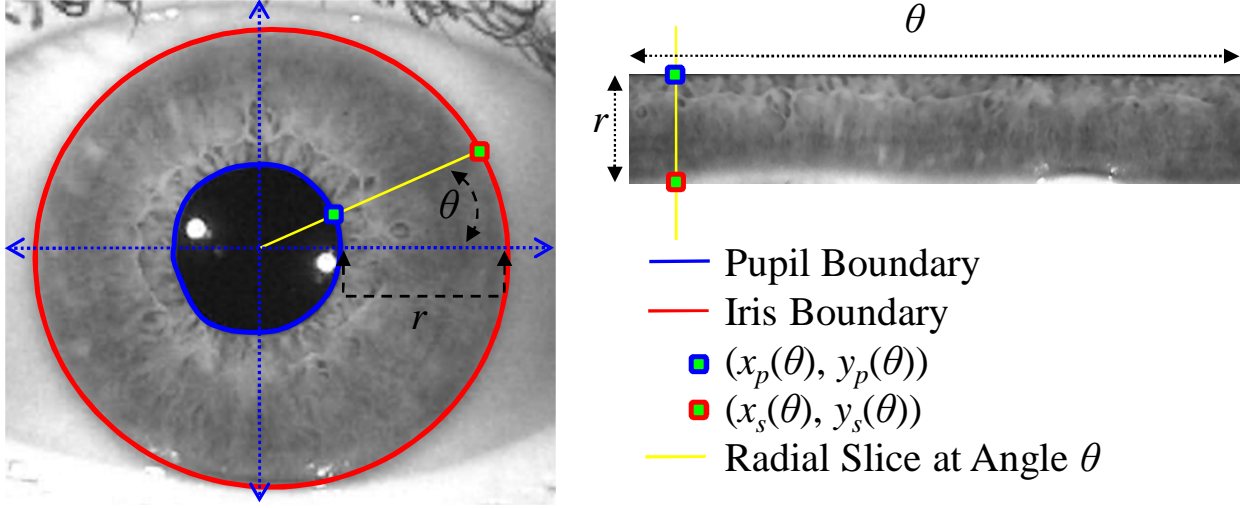


Figure 2.2: Illustration of the mapping between cartesian coordinates and polar coordinates. $(x_p(\theta), y_p(\theta))$ and $(x_s(\theta), y_s(\theta))$ represent a pair of points on the pupil and iris boundaries at angle θ and radius r respectively. This process effectively makes subsequent recognition steps robust to geometric distortions and linear deformations associated with pupil dilation and constriction.

code is supplemented with an equal number of masking bits, provided by the segmentation described in 2.1, which demarcates artifacts in the iris region, such as eyelashes, eyelids, and specular reflections.

The difference between two iris codes is calculated by measuring the dissimilarity between them, known as a fractional hamming distance. Artifact regions, demarcated by masking bits, are excluded from the computation. This is illustrated as follows:

$$HD = \frac{\sum_{i=0}^{2048} ((galleryCode_i \otimes probeCode_i) \cap galleryMask_i \cap probeMask_i)}{\sum_{i=0}^{2048} (galleryMask_i \cap probeMask_i)}, \quad (2.6)$$

where \otimes is the XOR operator and $galleryMask, probeMask$ represent the masks signifying which bits should be excluded from the comparison between $galleryCode, probeCode$ respec-

tively. This computation may be repeated several times while circular shifting either code along its angular coordinate (θ), keeping only the minimum hamming distance. This process aligns one code to the other which may be displaced due to roll rotation of the head. (Note that rotation of the eyeball along the yaw axis, resulting in an off-axis gaze or “off-angle” iris, will not be corrected by this computation. Instead, transformations that project an off-angle iris image back into frontal view have been considered [29, 34, 94].)

2.2.2 Wildes’ Framework

Wildes et al. [112, 114, 115] presented an alternative design to iris recognition while encompassing the same processing blocks as the Daugman system. Image acquisition under the Wildes et al. framework employs a diffuse light source and circular polarizers. This prevents specular reflection, created from the illumination source, from attenuating the iris region while diffuse lighting makes the acquisition process more comfortable to the human operator[†]. The system provides a reticle which aids the operator in positioning the eye for subsequent capture. Iris and pupil localization is realized via Hough transform on parametric definitions of a circle. First, the acquired image, $I(x, y)$, is smoothed with a Gaussian filter, $G(x, y)$. The gradient magnitude of the smoothed image is then thresholded to create a binary edge map.

$$E_{img} = |\nabla G(x, y) * I(x, y)|. \quad (2.7)$$

[†]The term operator in this context refers to the subject who’s iris is being imaged as opposed to an operator who is assisting the subject during the acquisition process.

Wildes et al.'s hough transform for a set of edge points, (x_j, y_j) is defined as:

$$H(x_c, y_c, r) = \sum_{j=1}^n h(x_j, y_j, x_c, y_c, r), \quad (2.8)$$

where

$$h(x_j, y_j, x_c, y_c, r) = \begin{cases} 1, & \text{if } g(x_j, y_j, x_c, y_c, r) = 0 \\ 0, & \text{otherwise} \end{cases}, \quad (2.9)$$

$$g(x_j, y_j, x_c, y_c, r) = (x_j - x_c)^2 + (y_j - y_c)^2 - r^2. \quad (2.10)$$

In particular, $h(x_j, y_j, x_c, y_c, r) = 1$, for every triple (x_c, y_c, r) , that describes a circle through

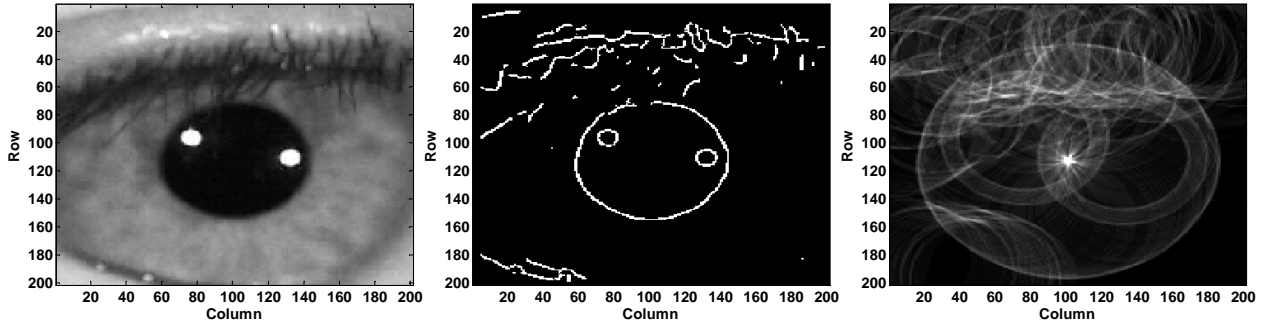


Figure 2.3: Illustration of circular Hough transform for finding the pupillary boundary. The input image (left column) is first smoothed, followed by thresholding of the gradient magnitude, in order to generate an edge map (center column). Finally, equations 2.8-2.10 are used to generate a Hough accumulator matrix (right column - Accumulator matrix generated with $r = 42$). In implementation, the search iterates over a range of radii and an accumulator matrix is generated for each radius.

edge point (x_j, y_j) . H is the accumulator array which stores all votes. This process is repeated for every edge point and the triple that maximizes H is used to represent the boundary. This process is repeated for the pupillary boundary but the parameter space is constrained to lie within the iris boundary. Eyelids are estimated in a similar fashion except that edge gen-

eration is biased towards horizontal edges while the parametrization of H is modified from circular to arcuate. The parameter search is constrained to be within the iris/sclera boundary but above or below the pupil boundary for upper and lower eyelids respectively.

Wildes et. al describe a representation based on multiscale Laplacian pyramids. Their empirical observations concluded that, “acceptable discrimination between iris images could be based on octave-wide bands computed at four different resolutions.” This representation is derived directly from the filtered image, yielding more information for the matching process but at the expense of requiring more resources in terms of storage with respect to Daugman’s representation. Registration between pairs of iris images is accounted for by finding a mapping function $(u(x, y), v(x, y))$ such that, for all (x, y) , the difference between $(x, y) - (u(x, y), v(x, y))$ in the input image is close to (x, y) in the model image. This ensures robustness to linear distortions between pairs of images such as translation, rotation, and scaling.

Matching pairs of irides is captured by spatially summing pixel differences within each frequency band between the input and template images. In practice, this is implemented by using normalized correlation over small 8x8 blocks in each frequency band. These correlation blocks are subsequently combined, to generate a similarity score for each frequency band, using the median statistic. An overall score is generated by combining the median correlation values from each frequency band through the application of Fishers Linear Discriminant, which minimizes the variance within a class of iris images while maximizing the variance between different classes.

2.3 Image Acquisition

Image acquisition, as described in section 2.2, is the process of capturing an iris image which is then subsequently processed for recognition. Standards such as the iris image data interchange [3] dictate that “acceptable” iris images must have at a minimum 150 pixels across the diameter of the iris. The gray levels separation between the iris and sclera should be at least 70 while the separation between the pupil and iris should be at least 50. With respect to subject presentation, eyes should be opened as wide as possible. The head should be positioned in the camera field of view such that the roll angle does not deviate the left and right iris centers by more than 10° . Eyeglasses (enrollment only), hard contact lenses, and patterned soft contact lenses should be removed. Furthermore, the eye is irradiated with near infrared illumination during acquisition, which in practice can be very challenging depending on the stand off distance. Traditional iris systems meet these requirements by imposing constraints on the user and the capture system (i.e. stand off distance, user cooperation, capture volume). The following is a summary of commercial acquisition systems followed by efforts leading to less invasive acquisition systems.

2.3.1 Commercial Acquisition Systems

The majority of commercial iris recognition systems require highly constrained subject presentation for successful operation. Table 2.1 provides a survey of selected commercial iris sensors along with their prerequisite operating parameters. Such systems employ near-infrared illumination (700-900nm) at a stand-off distance typically less than 2ft and acquire image data at a resolution of 480x640. Furthermore, substantial cooperation from the user

is required (i.e. remain stationary while gazing into the camera field of view. This process is usually supplemented with an audio or visual cue for precise eye positioning and focus).



Figure 2.4: Illustration of commercial iris acquisition systems. Top row: LG Iris Access 4000, Iris Guard AD100, Oki Irispass-M. Bottom row: Sarnoff Passport, Aoptix Insight, and GRI's HBOX-V. Note that systems on the bottom row impose fewer constraints on the user.

The aforementioned constraints, albeit intrusive to the user, ensure that the acquired iris image is of sufficient quality such that subsequent processing tasks perform with high fidelity. By contrast, the Aoptix Insight [102], Sarnoff Passport (also known as iris on the move [67]) and Glance [91], operate at greater distances with increased capture volume. Furthermore, the Passport does not require the user to remain stationary during acquisition; a normal walking pace will suffice without causing the system to fail.

[†]The Crossmatch I-SCAN 2 is a hand held device similar in operation to binoculars. The device is held up to the periocular region until an image is acquired.

Manufacturer	Model	Stand-off Distance	Acquisition Time
Crossmatch	I-SCAN 2	n/a [†]	2-15s
IrisGuard	IG-AD100	21-37cm	4s
LG Electronics	iCAM4000	26-36cm	2-4s
Oki	IRISPASS-M	30-60cm	< 60s
Panasonic	BM-ET330	30-40cm	< 60s
Less Intrusive Systems			
Aoptix	<i>Insight</i>	1.5-2.5m	4s
Sarnoff	<i>Passport</i>	3m	2s
Sarnoff	<i>Glance</i>	70cm	5s

Table 2.1: Survey of selected commercial iris recognition cameras. The majority of these sensors place constraints on the user for successful operation. The sensors in italics can operate at further distances or even while users are walking, resulting in a less intrusive experience.

Sensor Interoperability

The field of iris recognition continues to proliferate, not only in algorithm development (to accommodate the non-ideal trend), but also in sensing technology. As new sensors are produced, the question of interoperability arises. Are new sensors interoperable with existing sensors? This is a challenging problem for a number of reasons, not the least of which are related to the intended operational environment, sensor specific characteristics such as the underlying optics, IR wavelength, but also photometric covariates associated with non-ideal capture such as illumination.

Several studies in the field of fingerprint recognition [72, 87, 88], where sensors are based on a variety of technologies, have investigated this problem. With respect to iris recognition, several studies have explored the interoperability of iris images captured by the same sensor but at differing light wavelengths, more commonly known as multispectral iris recognition [12, 14, 126]. These works exploit the use of color information for improving iris

recognition performance. In [89], iris recognition beyond 900nm is investigated. The authors consider wavelengths within the short wave infrared band up to 1650nm.

In [105], the interoperability of three different iris capture systems were evaluated. Performance of each sensor was compared in a single sensor and cross sensor manner with three different matching algorithms. It was concluded that single sensor performance was observed to outperform cross sensor matching performance. However, no insight was provided on the understanding of imaging characteristics that impacted the cross sensor experiments. More recently, Connaughton et al. [22, 23] investigate the cross sensor performance of three different state-of-the-art iris sensors. They investigate how covariates such as changes in the acquisition environment and dilation ratio between iris images impact single sensor and cross sensor performance. They observed that changes in acquisition environment had the largest impact on recognition performance.

2.3.2 Unconstrained Subject Presentation

Recent research in iris image acquisition have focused on reducing constraints, leading to a less invasive experience for the end user. Mitsubishi electric research laboratories (MERL) [40] developed an automatic iris capture system which operates between 1.2m and 2.1m, captures both irides, and requires no interaction from the user. They employ a low resolution wide field of view (WFOV) video camera and a high resolution narrow field of view digital still camera. The video camera is used to detect faces while the digital still camera captures an image of both irides. Both sensors are mounted on a pan tilt unit (PTU), which allows the sensor system to move such that the eye region is registered with the still camera prior to capturing. A dual WFOV camera system was presented in [111], which can operate

up to a stand off distance of 1.5m. This system employs two WFOV cameras which are calibrated to ascertain distance, from camera to user, as well as direction. An iris camera is mounted on a PTU which moves to capture an image containing both left and right irides when the detection of a face is signaled from the WFOV cameras.

A more ambitious iris capture system, with respect to distance, was presented in [35]. Images were captured at distances up to 10m while using collimated near infrared illumination (880nm). Elliptical mirror telescopes were fabricated to facilitate image capture at 5m and 10m respectively. Each telescope was designed to capture an image with an average iris diameter of 128 pixels. However, the system was highly constrained with respect to user cooperation. The user was required to be seated while the user's head was supported with a chin rest and forehead stabilizer.

Matey et al. [67] pioneered an iris biometric system requiring only moderate user cooperation known as iris on the move (IOM). Their system affords increased capture volume, decreased acquisition time, increased capture distance, and the capability of imaging while users are on the move. The system also accommodates normal eye wear such as glasses and contact lenses. Users walk through an access control point, at normal walking pace, while looking forward. The system as described in [67] makes use of two vertically stacked COTS video cameras with strobed infrared illumination, capable of capturing an image with an average iris diameter of 100 pixels.

Eagle-Eyes [5,6] is a multi-modal biometric system capable of acquiring both face and iris at a distance through hierarchically-ordered fields of views. Their system is comprised of a wide area scene camera, a narrow field of view face camera, and dual iris camera. The fixed scene camera is utilized to detect and track humans which signals the face camera to capture

a higher resolution image of the face. This camera is mounted on a PTU with a rangefinder, laser illuminator, and dual iris camera. Eye tracking and range information, generated by the face camera and ranger finder, is used to adjust the zoom and focus of the iris camera lens to match the range trajectory of user. The system by design can accommodate moving targets as well as multiple users in the scene at the same time.

2.3.3 Iris Image Quality

The presence of heterogeneous image characteristics, such as blur, poor contrast, pupil dilation, deviation in gaze angle, and iris occlusion, are dictated by environmental and operational acquisition conditions. These characteristics are commonly referred to as “noise” in iris literature which can potentially confound subsequent processing iris processing stages. Typical mitigation strategies involve detecting these characteristics during acquisition and reacquire if the detection threshold is exceeded. Artifacts resulting from occlusion such as specular reflection, eyelids, eyelashes, and edges introduced from contact lenses are localized and subsequently excluded during template matching.

The effects of image blur make segmentation and matching more challenging by reducing edge contrast. Edges that distinguish the iris from surrounding image structures are lost or degraded, essentially making the entire image more homogenous. Daugman [27] demonstrated the use of a bandpass filter which measures the ratio of power in high frequency bands to slightly lower frequency bands. A similar approach was adopted by Kang [52]. Their proposed filter exploited finer frequencies than Daugman while reducing the size of the filter for increased computational efficiency. Ma et al. [64] exploited differences between blurred (defocus and motion) and non-blurred imagery in the Fourier spectra of local iris

regions. Zhang et al. [120] exploited the edge contrast of a local image region adjacent to the pupil.

Several authors consider multiple factors which may influence the performance of iris recognition systems. Proenca and Alexandre [83] identify five types of noise; eyelids, eyelashes, incorrect pupil localization, strong reflections, and weak reflections, which are observed to negatively impact performance. Subsequent processing estimates the aforementioned noisy characteristics based on local statistical features surrounding each iris pixel. A neural network is constructed based on these features in order to classify each input pixel as noise or noise free. Kalka et al. [50] describe measures for defocus blur, motion blur, illumination, off-angle view, occlusion, specular reflection, and pixel count. Each measure is combined through the application of Dempster-Shafer theory to generate a singular quality index for each input image. A thorough description of this algorithm is provided in Chapter 3. More recently, Daugman [29, 31] employs measures for characterizing image blur, interlacing, occlusion, and reliable iris code bits.

The amount of discriminating information within the iris varies locally along the radial and angular direction of iris tissue. Several authors exploit this observation for generating a quality measure. Chen et al. [19] quantify the energy response from multi-scale 2D-Mexican hat wavelets along concentric bands of iris tissue increasing radially from the pupil. Each band is weighted such that bands closer to the pupil (e.g. collarette region) provide more weight. Belcher et al. [7] propose an information distance measure, which compares feature information to that of a uniform distribution, as a means to quantify the information content of iris features. In practice, the gradient magnitude is computed with a sliding window along the polar representation of the iris. Monotonically decreasing weights are introduced such

that windows closer to the pupil boundary apply more weight than windows near the outer iris boundary.

In [100], the NIST Iris Quality Calibration (IREX-IQCE) evaluation focused on identifying iris image quality components that are algorithmic and/or camera agnostic, and are influential on the performance of iris recognition system. Quality algorithms were voluntarily submitted by 14 vendors. Analysis was conducted across three iris data sets having a broad range of quality, ICE 2006, OPS and QFIRE. The following iris image characteristics were evaluated: useable iris area, contrast between the iris and pupil boundary, pupil boundary shape, contrast between the iris sclera boundary, gaze angle, sharpness, dilation, interlace, gray scale spread, iris shape, motion blur and signal to noise ratio. Of these characteristics, useable iris area was observed to have the most significant impact on iris matching performance, followed by iris/pupil contrast, shape of the pupillary boundary, iris/sclera contrast, gaze angle and sharpness. It is also important to note that this evaluation included several novel morphological characteristics (e.g., shape of the iris or pupil boundary), which have not been studied in prior approaches as a characteristic for iris image quality. The reader is referred to [100] for more thorough discussion and details.

The research presented in this section can be dichotomized into two categories: global and local analysis. Global image assessment is performed on the entire image and does not use intrinsic information specifically pertaining to the iris. Alternatively, local analysis uses only information related to the iris which requires iris and/or pupil segmentation. Global analysis is typically computationally faster because it does not involve segmentation. However, it is much more difficult, if not impossible, to get an accurate assessment of the iris through global analysis because heterogeneous image structures negatively influence the as-

assessment. On the other hand, local assessment can provide a much more accurate analysis of the iris but only with accurate segmentation.

2.4 Segmentation

Segmentation is the process of isolating iris tissue from surrounding periocular image structures. It is arguably the most important processing task because errors introduced at this stage will propagate through the rest of the system, negatively impacting performance. Traditional means of locating these boundaries were discussed in subsections 2.2.1 and 2.2.2. Variations, alternatives, and improvements of these methodologies have been proposed in the literature. The following is a selected summary of efforts on iris segmentation research.

2.4.1 Orthogonal View Processing

Huang et al. [47] employ a combination of the circular based Hough transform and the integro-differential operator. The input image is first downsampled to reduce complexity followed by canny edge detection and thresholding to generate a binary edge map. Parameters of the circles, representing the coarse iris and pupil boundary, are recovered from the edge map using a Hough transform. Subsequent use of the integro-differential operator refines the parameters for each boundary. Eyelids are estimated as in [112] while eyelashes are masked out through use of hard thresholding. Tian et al. [103] make use of a “self adapting window” to coarsely locate the centers of the pupil and the iris from a binary image. Subsequent use of a Hough transform, which is constrained to the region of the self adapting window,

yields a refined estimate of the circle parameters for both boundaries while at the same time increasing the efficiency of the Hough transform.

Camus and Wildes [15] employ a variation of the integro-differential operator for estimating the pupil and iris boundary. First, specular reflections are scrubbed and local minima are selected, in 5x5 pixel blocks by hard thresholding, as seed points. The local image region, surrounding each seed point, is transformed to polar representation with the seed point as the center. A rough estimate for the pupil radius is recovered by iterating through radius values that maximize a “goodness of fit” measure. This metric is based on local derivatives in the polar representation from eight radial directions, that are tuned to estimate gradient strength, uniformity, and bias toward dark regions on the boundary interior. A coarse iris boundary is estimated in a similar fashion, except that the boundary is constrained to lie outside the estimated pupil boundary. The sum of both pupil and iris measures is used as a local “goodness of fit” for each seed candidate. The maximum across all seed candidates is chosen for the correct boundaries.

Bonney et al. [9] locate the pupil edge region by converting the input iris image to its least significant bit plane. This is followed by the use of morphological operators, erosion and dilation, until a single homogeneous region remains. The end points of this region are used for fitting an ellipse. To isolate the iris boundary, local standard deviation windows are thresholded in the vertical and horizontal direction to create a binary mask. A “guess and check” method is used in combination with the mask to find the end points for all cardinal directions. An ellipse is fit to the recovered end points.

In [60], Liu et al. focused on improving the publicly available iris segmentation provided by Masek [65] which utilized the Hough transform. The first improvement is based on obser-

variations from near infrared iris imagery. The contrast between the pupil and iris is typically more pronounced than that of the iris and sclera. Therefore, the detection order was reversed. With a more reliable estimate of the pupil, errors introduced from constraining the search of the iris are reduced. In order to reduce false iris boundary candidates in Hough space, the edge map was thresholded to reduce spurious edges introduced from specular reflection and iris texture. Furthermore, votes for center locations are constrained to 30° on each side of an edge's normal direction. The last improvement is a set of boundary constraints which ensure that the interior region of a boundary is always darker than the exterior region.

2.4.2 Non-Orthogonal View

Acquisition systems that operate in less constrained environments, such as those illustrated in subsection 2.3.2, allow much more variation in the images that are captured in contrast to traditional systems. Indeed, an important area of research in iris segmentation is specifically focused on processing images captured from non-orthogonal view. Such imagery is characterized by motion/defocus blur, non-uniform illumination, heavy occlusion, and variation in gaze angles which negatively impacts system performance [68].

One of the main limitations with the Hough transform is that is highly dependent upon the input edge map, which in itself is dependent upon local image characteristics such brightness, contrast, and extant noise. Traditional acquisition systems place constraints on the user which attempt to minimize unfavorable imaging characteristics. By contrast, when processing non-orthogonal view iris images, noisy characteristics are more prominent, requiring a robust methodology for edge generation. Proenca and Alexandre [82] tackle this

problem by generating an edge map that is less dependant upon local image characteristics. They extract three features from the input iris image, image locality (x,y) and intensity, and cluster the input image into three regions using a fuzzy k-means clustering algorithm. This facilitates the generation of a homogeneous edge map with minor influence from local image characteristics. Subsequent use of the Hough transform recovers parameters for the circles representing the pupil and iris boundaries. More recently, Proenca [81] provided an improved clustering methodology utilizing multilayered perceptron feedforward neural networks. Classification is performed in two stages. First, a network is employed to classify the image into two regions, sclera and non-sclera, using central moments derived from hue, blue, and red chroma color components. A second network is used to distinguish the iris from surrounding regions. This network uses features derived from local image saturation, blue color component, and a novel feature called “proportion of sclera” which is generated from the output of the sclera network. Finally, boundaries are fit to the resulting mask, representing the classified iris and pupil pixels, with a 15th degree polynomial.

Tan et al. [101] present a coarse-to-fine strategy that focuses on improving the speed and accuracy of Daugman’s integro-differential operator when processing visible wavelength iris data captured under unfavorable conditions. They argue that periocular image structures with excessive local contrast confound traditional segmentation methodologies. In order to avoid this, a coarse classification is employed to delineate iris and non-iris regions using the red component of the visible wavelength image. This is achieved by clustering image pixels (iris, eyebrow, skin, glasses) through the combination of semantic priors and local 8-neighborhood image statistics. An integro-differential constellation operator is utilized to estimate the inner and outer iris boundaries from coarse localization. This operator is four

times faster than the traditional operator, attributed to the combination of a greedy search algorithm and predefined search path or constellation. Next, the shape of both boundaries is refined via intensity statistics to account for irregularities when the actual boundaries are not circular. Eyelids are estimated by parabolic Hough transform after horizontal rank filtering the eyelid region. Eyelashes are estimated in similar manner to [29].

The shape of the iris inner and outer boundaries have traditionally been modeled as circles and/or ellipses despite observations suggesting the appropriateness of a more complex shape. Indeed, it is not uncommon for these boundaries to have an irregular shape which is not captured by a circle or ellipse. Daugman [29] presented a new approach for isolating these boundaries based on active contours. This approach allows for a flexible contour while at the same time providing robustness to discontinuities, resulting from eyelids or eyelashes, in the iris region. The actual shape is approximated by a discrete Fourier series expansion of the contour data which inherently provides control over the complexity of the shape and the degree of smoothness. This representation is further motivated by the fact that geometric distortion introduced from variation in gaze angle is captured within the Fourier expansion of the contour data itself. This information, once estimated, is used to transform the off-angle iris image back into frontal view as in [34].

Shah and Ross [97] present an approach to model the outer iris boundary based on level sets, referred to as geodesic active contours (GAC). The basic premise is to evolve a curve, which starts on the exterior of the inner (pupil) boundary, until the outer iris boundary is delineated. Curve evolution is modulated by a set of constraints that avoid over-segmentation by minimizing the thin plate spline energy while ensuring smoothness. Points on the final curve are chosen to estimate a circle which is the final representation for the outer boundary.

2.4.3 Segmentation Evaluation

Errors introduced during segmentation propagate through subsequent stages of iris processing, specifically normalization and feature extraction, confounding system performance. Therefore, it would be advantageous to automatically detect such errors if they occurred. A relatively new avenue of research in iris literature is focused on evaluating the outcome of the segmentation block.

Lee [57] assumes that the gray level intensities within the pupil and iris regions are relatively uniform. Therefore, a favorable segmentation should maximize the uniformity within each region and maximize the contrast between different regions. This notion is captured by measuring intra-region homogeneity and inter-region heterogeneity; the former characterizes region uniformity while the latter characterizes differences between distinct regions. Lee employs eight distinct gradient operators to characterize the homogeneity and heterogeneity of the pupil and iris region. The estimated measures are combined through the application of PCA to generate a single segmentation prediction index. A similar approach was illustrated by Zhou et al. [123] using local homogeneity measured from four regions along the iris boundary.

Zuo and Schmid [127] describe an approach which automatically evaluates the precision of iris segmentation. They argue that the gradient should be strong along the inner and outer iris boundaries for favorable segmentations. Therefore, a suitable descriptor for characterizing segmentation precision, should be a function of gradient strength along both contours while excluding regions which may be attenuated by occlusion. In practice, the gradient is derived from the polar representation of the iris along the angular variable. This measure is supplemented with a global constraint based on region intensity such that interior boundary

regions should be darker than exterior.

Kalka and Bartlow [49] utilize a combination of geometric and intensity features to discern whether an estimated segmentation is favorable or not. Given a segmentation, the pixels within the pupil boundary are classified as pupil or non-pupil pixels using a likelihood ratio test. This process is repeated while increasing the radius of the estimated pupil. The former metric captures information about over-segmentation while the latter can be used to characterize under segmentation. Both features are supplemented with a geometric feature that corresponds to the distance between the estimated pupil and iris center. In practice, the pupil and iris are not concentric but their centers are relatively within close proximity of each other. Therefore, the distance between both centers can be used as a simple feature to predict gross inaccuracies in the iris boundary. An overall evaluation result is generated through the application of machine learning on the three aforementioned features (see Chapter 4 for a thorough description and updates to this algorithm).

Chapter 3

Iris Image quality

3.1 Introduction

In this chapter we introduce a comprehensive approach to assess quality from an iris biometric image. We identify a broad range of factors including defocus blur, off-angle, occlusion/specular reflection, lighting, and iris resolution. We then analyze their effect on traditional iris recognition systems. Publicly available iris data sets such as CASIA v3.0 [1], and ICE 1.0 [61] offer images with varying quality factors. The West Virginia University (WVU) data sets, also publicly available, have a broad range of quality factors present in images. West Virginia University (WVU) non-ideal [25], and West Virginia University off-angle (WVUOA) [2] are non-ideal iris image data sets. WVUOA is an off-angle data set collected in the NIR spectrum. This data set consists of only off-angle data which we utilize in the off-angle experiments in this paper. In order to get a data set with a broad range of factors we systematically collected an iris image data set of varying quality affected by defocus, occlusion/specular reflection, off-angle, lighting, and iris resolution. The intent

of this analysis is to evaluate how these factors influence recognition performance and to what extent. We are also aware of the availability of UBIRIS [80], the data set collected in the visible light spectrum. Traditional / commercial iris recognition systems operate in the NIR spectrum, typically 700 - 900nm. NIR imaging is used because it helps emphasize iris texture which can be difficult to image under visible light (i.e. texture is much richer for darkly pigmented eyes under NIR compared to visible). The main focus of this paper is to study and measure the impact of iris quality on traditional/commercial systems which utilize NIR imaging. Therefore, studying the quality factors in UBIRIS images is outside of the scope of this paper. Additionally, NIR is used because the human eye is not as sensitive to NIR wavelengths as to visible light [27]. The pupil and eye remain open under NIR illumination. However, a number of precautions has to be taken to ensure eye safety when dealing with NIR illumination. This is of particular concern in unconstrained environments when stronger NIR sources are required. The International standard for laser safety and equipment [21] should be consulted when designing collection experiments. Exposure to wavelengths between 700 – 1050 should not exceed an irradiance of $10^{.0002(\lambda-700)} W m^{-2}$ for captures lasting between 100 – 1000 seconds.

Next we design procedures for estimation of defocus blur, motion blur, occlusion, specular reflection, lighting, off-angle, pixel-counts, and apply these to the images from data sets mentioned above. The individual factors are then “combined” using a fusion algorithm. Although we evaluated several evidential reasoning fusion approaches [71], here we present the Dempster-Shafer criterion [73, 74, 96] which offers several advantages for combining the image quality scores. Fusing various iris quality attributes into a single “score” is somewhat controversial, yet it is practically important. Skeptics argue that the informative value of

a single fused iris quality score diminishes through fusion. For example, there is no doubt that if the goal of quality measurement is to improve the collection protocol, individual quality factors are more informative. On the other hand, vendors of biometric systems as well as the users desire to have a unified quality score as a measure of overall suitability for authenticating an individual. Further, recent “quality-enhanced” multi-modal biometric fusion algorithms [70, 75] tend to use a single quality estimate for each biometric modality included in the fusion framework. Our work offers clear contributions in both of these contexts: (1) Scores that describe specific iris image quality attributes can be easily used for improving collection protocol. In fact, in this paper we compare the distribution of quality factors in each of the data sets listed above, thus offering an independent insight into the collection conditions. (2) We also demonstrate that the fused iris quality score clearly correlates with the matching score, which indicates that it can be used as an informal match confidence measure.

3.2 Non-ideal data set collection and Experimentation

We collected a total of 300 images from ten individuals comprising a total of twenty classes. The images were collected using a Hitachi KP-F120F monochrome CCD camera in the NIR spectrum utilizing a Navitar 7000 lens at a resolution of 1280x960. This model captures in single frame and/or video format modes. We used two LED illuminators from a sony DSC-F717 that have a peak wavelength of 850nm. In order to filter out visible light we utilized a B+W 093 infrared filter.

From each subject we collected three ideal, defocused, occluded, reduced lighting, and

reduced resolution images. An adjustable tri-pod with a camera clasp was used as a “chin plate” so subjects could rest their heads while imaging to reduce involuntary movement of the head. Because of the non-ideal nature of the data we utilized an in house iris segmentation algorithm developed by Zuo et al. [125] designed specifically for non-ideal iris. To evaluate the influence of individual quality factors on recognition performance we invoke Libor masek’s [66] publicly available Log-Gabor filter based iris encoding algorithm (default parameters used). The corresponding metric that is used as a measure of performance is Hamming distance. To demonstrate the dependence of verification performance on individual quality factors, we plot Hamming distances as a function of the degradation level with respect to each quality factor. Each figure demonstrating the degradation of performance will contain two plots: an error-bar (standard deviation) plot of genuine scores and an error-bar plot of imposter scores displayed as functions of a parameter characterizing the strength of a quality factor under study (for all graphs the first strength value represents scoring of the ideal imagery). For each subsequent level the ideal imagery is used as the gallery images and are compared to the degraded imagery. Listed below are procedures that were carried out to collect and evaluate the quality factors.

3.2.1 Defocus Blur

Defocus blur can result from many sources, but in general, defocus occurs when the focal point is outside the “depth of field” of the object to be captured. The further an object is from depth of field interval the higher the degree of defocus. In order to collect varying levels of defocused images, subjects were initially positioned six inches from the lens with their chins lined up at the edge of the chin rest. This was the initial position that the ideal

images were collected at (the camera was manually focused in at this position). Subjects were then instructed to slide their head back along the chin rest and frames were recorded at approximately $\frac{1}{8}$, $\frac{1}{4}$, and $\frac{3}{4}$ inches from the edge of the chin rest.

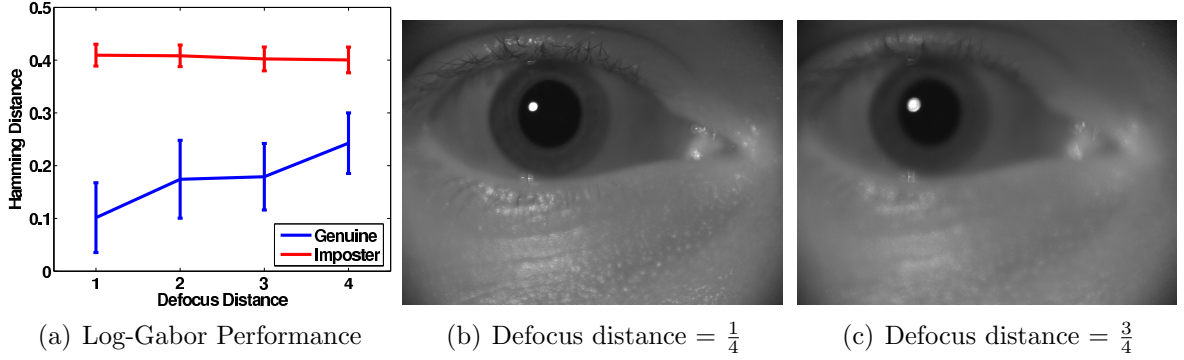


Figure 3.1: (a) Effect of defocus blur on Log-Gabor encoding technique. Distance 1 corresponds to the ideal images and 2, 3, and 4 represent when the subject is $\frac{1}{8}$, $\frac{1}{4}$, and $\frac{3}{4}$ inches from the edge of the chin rest respectively. (b) & (c) are sample images taken from $\frac{1}{4}$ and $\frac{3}{4}$ inches from the edge of the chin rest.

Fig. 3.1 (a) is an error bar plot of the mean hamming distances. Defocus distance corresponds to the distance the subject is from the edge of the chin rest. Figure 3.1 (b) and (c) represent samples images from a subject at $\frac{1}{4}$ and $\frac{3}{4}$ inches from the edge of the chin rest. The main degradation is seen in the genuine scores as they increase the further the subject is from the edge of the chin rest. The imposter scores on the other hand remain virtually unaffected as the distance increases.

3.2.2 Off-Angle

Iris images which are not of frontal view are of special interest. For evaluating the effect of off-angle on performance, testing is done using the WVU off-angle iris image database. The

database has 208 iris classes, four images per class including two frontal view images, one from 15 degree view image, and one 30 degree view image. The initial angle values are those assigned during the data collection. To evaluate the effect of off-angle, we train recognition systems on frontal view images and test on off-angle images. The dependence of matching

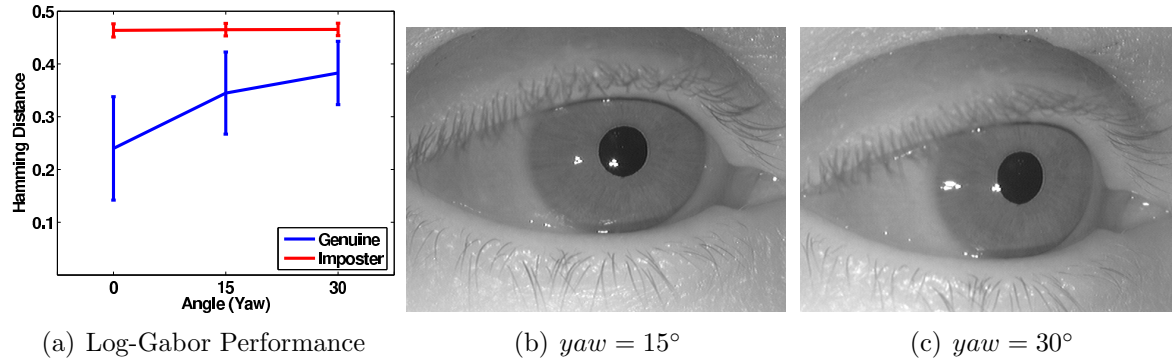


Figure 3.2: (a) Performance of Log-Gabor filter-based recognition algorithm vs. increasing yaw angle. (b) and (c) are sample images from the data set.

score values on the angle is displayed in Fig. 3.2 (a). As the yaw angle increases, the relative distance between the imposter and genuine score curves decreases. As with defocus, the main degradation is seen in the genuine scores.

3.2.3 Occlusion

Occlusion results from eyelashes, eyelids, camera orientation, hair, eye glasses, printed contact lenses and specular reflections that obscure iris texture. To compensate for the effect of occlusion, traditional iris recognition systems mask out the occluded iris regions. This quality factor was collected as video while subjects were instructed to open and close their eyelids. Similar frames from all subjects were pulled out from these videos based on the occlusion estimation technique described in the next section and used in the following ex-

perimentation. Each level was chosen such that successive levels linearly contained more occlusion than the previous. To estimate the amount of occlusion at each level, we used the masks obtained from our non-ideal segmentation. The ideal images represented in level 1 had a mean occlusion score of 0.23. Levels 2, 3, and 4 had mean occlusion scores of 0.34, 0.42, and 0.54 respectively. To explore how occlusion impacts performance, occlusion masks are only utilized for the ideal gallery images. Masks are not used for the actual occluded imagery. Fig. 3.3 reflects the impact of occlusion on recognition performance for the Log-

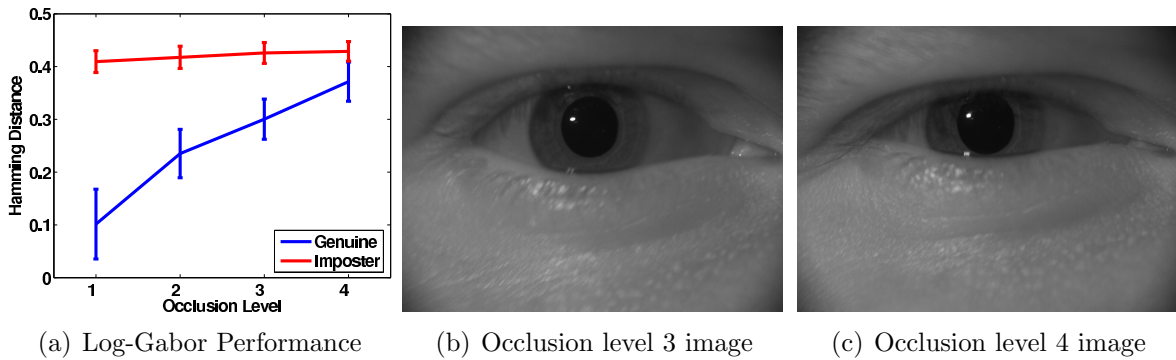


Figure 3.3: (a) Degradation of recognition performance due to varying levels of occlusion. (b) and (c) represent occluded images from a subject for the level 3 and level 4 groups respectively.

Gabor filter based encoding technique. As the amount of occlusion increases the separation between genuine and imposter scores decreases with the main degradation seen in genuine scores.

3.2.4 Lighting Variation

Traditional iris recognition systems operate at close range, less than 0.5 meters and subjects are stationary. This constraint typically allows for uniform lighting across the iris with sufficient brightness that emphasizes the rich structure of the iris. However, if the subject is

uncooperative or hasn't acclimated to the capture system non-uniform lighting patterns can occur (see Fig. 3.4). Furthermore, when distance and movement constraints are relaxed,

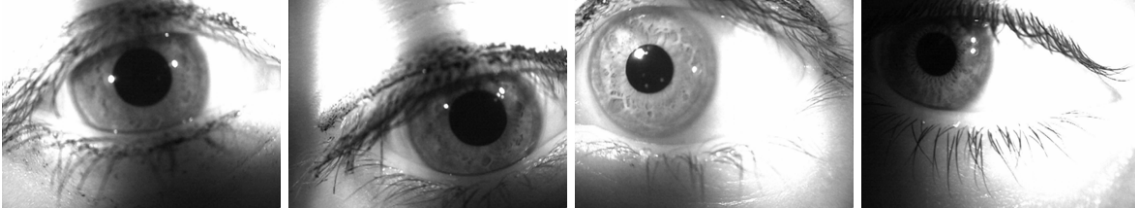


Figure 3.4: Example of images with non-uniform lighting conditions from the WVU data set.

such as in surveillance or port of entry type applications lighting issues such as deficiency can occur. In [69], the Iris On the Move (IOM) system requires a modest level of user cooperation to function properly. If the subject's walking pace is faster than 1m/s, captures could potentially suffer from insufficient illumination. The authors also note that after 3 attempts per subject the cumulative recognition rate is 100%. This result implies that subjects acclimate to the capturing system by their third attempt.

For this quality factor, we look at how lighting deficiency effects recognition performance by moving our NIR light source at varying distances from the subject. While we are interested in studying all forms of lighting variation, our current NIR light source is not capable of producing realistic non-uniform variation such that we can methodologically collect images without introducing long collection times for our subjects. To this end, we collect images with our light source positioned 1 (our ideal setting), 2, 3, and 4 feet behind the camera. Fig. 3.5 (a) is a plot of the performance of these images collected with the light source at the specified distances. As the distance increases the genuine scores increase.

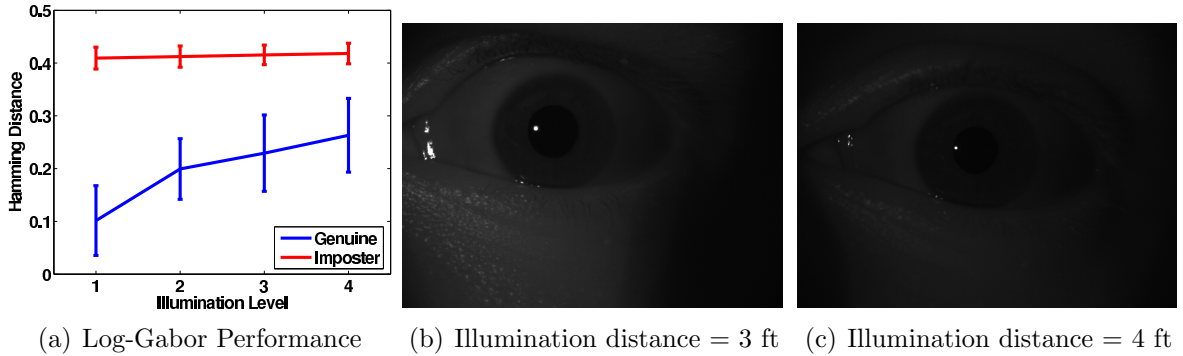
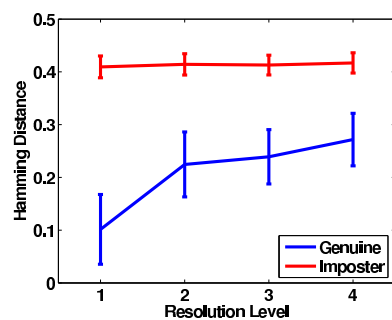


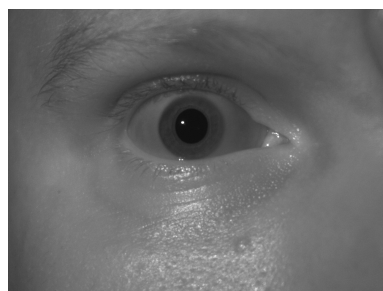
Figure 3.5: (a) Influence of lighting deficiency on Log-Gabor based encoding. (b) and (c) represent sample images from a subject when the light source is 3 and 4 feet behind the camera respectively.

3.2.5 Iris Resolution

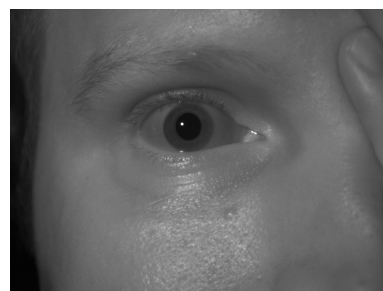
The amount of information within an iris is a function of its resolution. Capturing iris at a distance can result in variable resolutions. With that notion in mind, at which point does the information contained within the iris become unsuitable for distinguishing individuals based on iris patterns? The ISO iris image standard [38] states that 200 pixels across the iris is considered good quality for recognition. We wanted to explore lower resolutions since iris at a distance is gaining much more interest and capturing high resolution images of the iris can be problematic. To this end, we collected iris data at four different resolutions. Our ideal data has an average iris diameter of 415.27 pixels while levels 2, 3, and 4 have an average iris diameter of 223.00, 156.89, and 124.11 respectively. Fig. 3.6 (a) illustrates the performance of Log-Gabor encoding while varying the resolution of the iris. As the diameter of the iris decreases the separation between genuine and imposter scores decreases.



(a) Log-Gabor Performance



(b) Iris diameter = 223



(c) Iris diameter = 168



(d) Iris diameter = 120

Figure 3.6: (a) Dependence of recognition performance on iris diameter. (b), (c), (d) are images from a subject with iris diameter of 223, 168, and 120 pixels respectively.

3.2.6 Discussion

The purpose of these experiments was to analyze the effect of various quality factors such as defocus, off-angle, occlusion, lighting, and resolution on the Log-Gabor based encoding technique in order to understand how or if these quality factors influence recognition performance. While we can not say which factors are more influential than others since their levels of degradations are of different scales we can at least say that all factors in this experiments negatively influence performance. The major impact of image degradation specifically influenced the genuine match scores for factors studied, resulting in a decrease in separation between the distributions of genuine and imposter match scores.

3.3 Estimation of Individual Quality Factors

In our initial work [51], we employed our own implementation of Wildes’s [113] algorithm using a “rough segmentation” that was based on image downsampling. This method was chosen because it provided better results than other segmentation algorithms that were available. However, this algorithm still resulted in substantial segmentation failures when evaluating non-ideal data sets such as WVU. In order to reduce these failures we adopted the same segmentation algorithm that was utilized in the data collection experiments which was specifically designed for non-ideal data.

3.3.1 Defocus

Defocus primarily attenuates high spatial frequencies. Due to this relationship, defocus can be assessed by measuring high frequency content in the overall image or segmented iris region. Daugman demonstrated this in [28] by proposing an (8x8) convolution kernel (table 3.1) and measuring the total power in the response. This 2-D spectral power is then passed through a compressive non-linearity of the form:

$$f(x) = 100 \frac{x^2}{(x^2 + c^2)} \quad (3.1)$$

in order to get a normalized score between [0, 100]. Here x is the total spectral power measured by the (8x8) convolution kernel and c is the half-power of a focus score corresponding to 50%. This spectral measure of focus works well when iris images are dominated by the iris as in Fig. 3.7 (a). However, when dealing with imaging not dominated by the iris, this spectral measure of focus can be misleading as in Fig. 3.7 (c).

-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	3	3	3	3	-1	-1
-1	-1	3	3	3	3	-1	-1
-1	-1	3	3	3	3	-1	-1
-1	-1	3	3	3	3	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1

Table 3.1: Daugman's 8x8 convolution kernel

Fig. 3.7 (b) and (d) display the responses of the band pass filtering when applied to Fig. 3.7 (a) and (c), respectively. Notice in Fig. 3.7 (b) that the iris region contains a significant amount of high spatial frequencies hence a highly focused image. In Fig. 3.7 (d) on the other

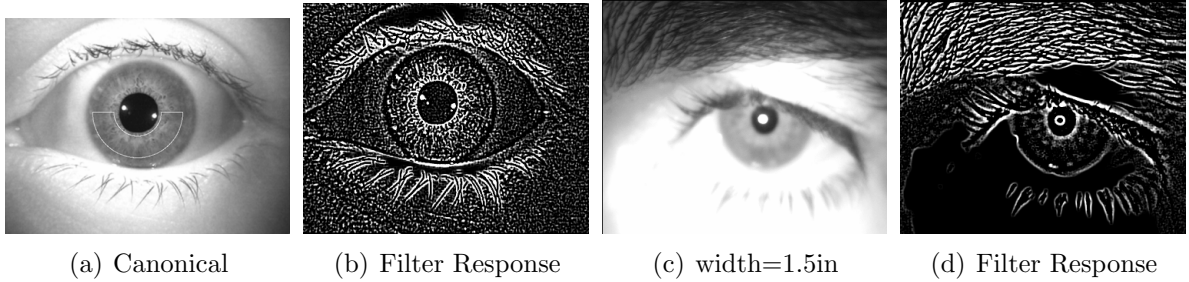


Figure 3.7: Sample images from WVU data set and their band-pass filter responses. The local assessment in (a) is performed based on the pupil and iris localization. The total width of the assessment region is $\frac{1}{3}$ of the iris region. The assessment region begins from $\frac{1}{7}$ of the distance from the pupil to iris boundary.

hand, the iris region does not contain high frequency information. We conclude therefore that the iris region is defocused. However, introduction of in focus eyebrows in (d) results in a high focus global score hence the need for local focus assessment. To compensate for this we employ the same spectral measurement of focus but locally rather than globally. We chose the bottom region of the iris as illustrated in Fig. 3.7 (a). This region was experimentally chosen as the upper iris region is more likely to be occluded from the upper eyelid. Image (a) has a defocus score 0.01 and (c) has a defocus score of 0.99 when using local assessment.

3.3.2 Motion

Motion blur results from the relative motion between the object or camera during exposure time, which can lead to linear and non-linear blur. Currently our work only includes estimation of linear motion.

Estimating linear motion blur is equivalent to estimating the primary direction of blur in the image, along with the strength of blur in this direction. To estimate the angle, we apply

directional filters in Fourier space. The input image is subjected to a Fourier transform as seen in Fig. 3.8 (arrow A). The dot product between the log-magnitude of the transformed input image and directional masks/filters similar to those shown in Fig. 3.8 (arrow B) (at 36 equally spaced orientations in the range (0, 180) degrees) is performed. The total power is calculated from each of these responses. The response with most directional power gives an estimate of the angle as seen in Fig. 3.8 (arrow C). Note that in this example the original image was blurred in the direction of 45° with a strength of 8 pixels. Let I be the image. Denote by $F(I)$ the Fourier transform of I . To find the estimate of the motion blur angle, we apply directional filters of a given scale α . Denote the filter response at an orientation Θ by $H(\Theta:\alpha)$. The following equation describes the procedure of estimating the direction of blur:

$$\hat{\Theta} = \arg \max_{\Theta \in [0:5:180]} \|F(I)H(\Theta : \alpha)\|^2. \quad (3.2)$$

Once the orientation of blur is obtained, strength is estimated by analyzing a slice of the log-magnitude image perpendicular to the estimated angle of motion blur and passing through the center. Note the main “lobe” of the Fourier spectrum in Fig. 3.8, above arrow B. The width of this main “lobe” is inversely proportional to the amount of motion blur strength. By measuring the power in the main “lobe” we can obtain an estimate of linear motion blur strength. In order to locate this lobe, the coefficients are first smoothed by using a B-spline with least-squares optimization (shown by arrow D). Finally a gradient based approach is adopted to locate the central peak (shown in arrow D). An example of a log-magnitude slice before and after B-spline smoothing is shown in Fig. 3.9.

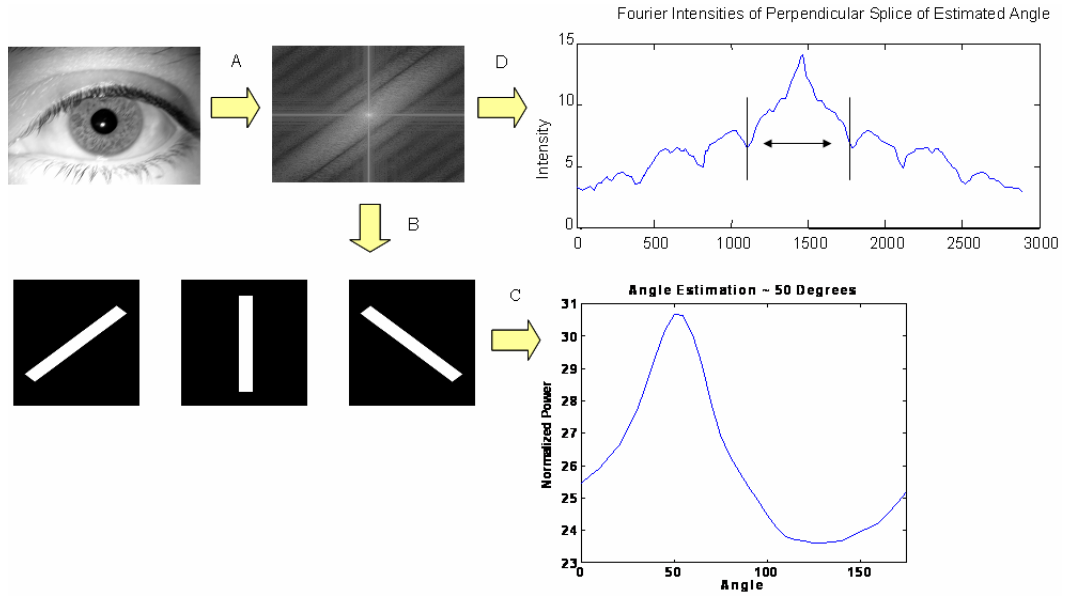


Figure 3.8: Motion estimation block diagram

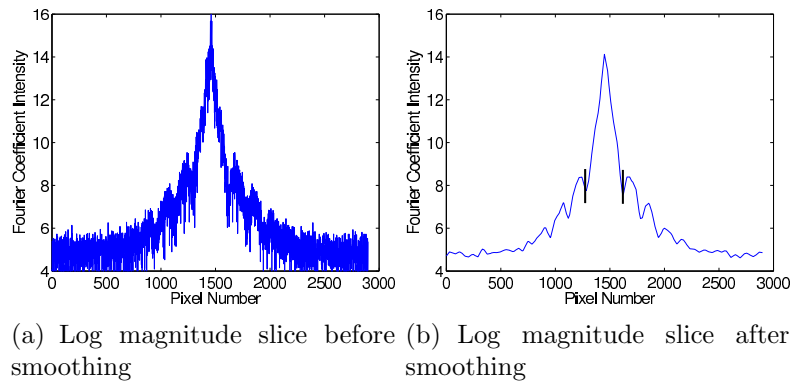


Figure 3.9: Log magnitude slice before and after b-spline smoothing.

3.3.3 Off-Angle

To estimate the gaze direction we assume that a rough initial estimate of the angle is available. To be more specific, we use the angle assigned during data collection as an initial estimate. We consider the case when a single iris image is available for estimation of the gaze direction and use circularity of pupil as a measure of the off-angle. The objective function applied to measure the pupil circularity is Daugman's integro-differential operator. We pick the estimates that maximize the value of the integro-differential operator. To be more specific, let ψ and φ be a pitch and a yaw angles, respectively. Let $J(\psi, \varphi)$ be an objective function that has to be optimized. For each given value of ψ in the range $\psi \in [\psi_{min}, \psi_{max}]$ and for each given angle of φ in the range $\varphi \in [\varphi_{min}, \varphi_{max}]$, (i) the off-angle iris image is rotated using the projective transformation and (ii) the objective function $J(\psi, \varphi)$ is calculated. Let $A(\psi, \varphi)$ be a 3×3 projective matrix parameterized by pitch and yaw angles. Let $I(x, y), x = 1, \dots, N, y = 1, \dots, M$, be an iris image. For each given value of $\psi \in [\psi_{min}, \psi_{max}]$ and $\varphi \in [\varphi_{min}, \varphi_{max}]$, the projective transformation is performed with respect to the pupil center, that is, each pixel (x, y) in $I(x, y)$ is mapped into a pixel $(x', y') = (x, y)A(\psi, \varphi)$ of the transformed image $I(x', y' : \psi, \varphi)$, where colon indicates dependence on the angles ψ and φ . Then the estimate of the angles $(\hat{\psi}, \hat{\varphi})$ solves the following optimization problem:

$$(\hat{\psi}, \hat{\varphi}) = \arg \max_{\psi, \varphi} J(\psi, \varphi), \quad (3.3)$$

where

$$J(\psi, \varphi) = \max_{r, x_0, y_0} |G_\sigma(r) * \frac{\partial}{\partial r} \oint_c \frac{I(x', y' : \psi, \varphi)}{2\pi r} ds|. \quad (3.4)$$

In our experiments we use bilinear interpolation to obtain the transformed image. Once the

angles are estimated we apply projective transformation [24] to rotate the off-angle image into a frontal view image. Fig. 3.10 is an illustration of the off-angle estimation.

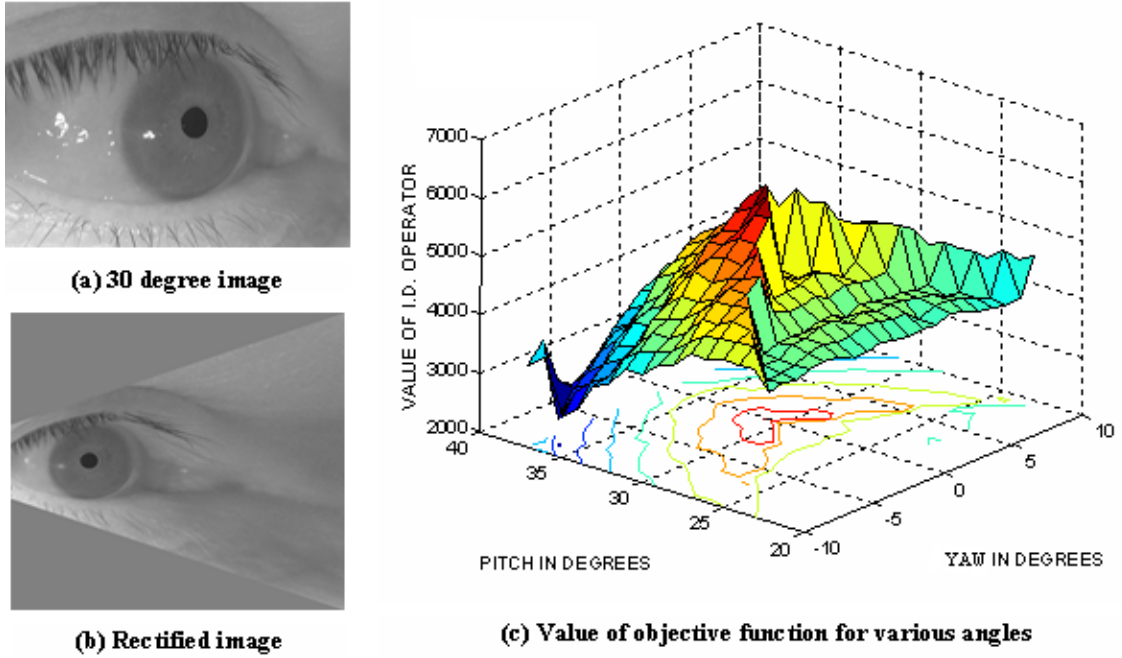


Figure 3.10: Off-Angle Estimation (a) An iris image that is out of plane by 30° which is assigned during data collection. (b) The image in (a) brought back in to frontal view by projective transformation. (c) A plot of the integro-differential operator over various pitch and yaw angles.

3.3.4 Occlusion

Occlusion, specular reflection and lighting variation are estimated sequentially. This is to reduce spurious measurements introduced from eyelid occlusion, when estimating specular reflection. In order to estimate light variation accurately it is prudent to remove specular reflections since the resulting influence will acutely impact lighting variation. With regard to that, occlusion is measured first. Our metric for occlusion has been extended based on

our earlier work [125] on non-ideal iris segmentation.

1. First, horizontal edges are detected by taking the gradient in the x direction, $\frac{\partial I_{x,y}}{\partial x}$ (Fig. 3.11 (b)).
2. The horizontal edges are “extended” using a morphological dilation [95] with a cross-shaped neighborhood (Fig. 3.11 (c)).
3. The “extended” edge map is then overlayed with the iris localization (Fig. 3.11 (d)).
4. All pixels not belonging to the “mainland” centered at the pupil are removed (Fig. 3.11 (e)).
5. Finally any discontinuities in the mask that remain are filled (Fig. 3.11 (f) and the percentage of masked pixels is used as our measure for occlusion).

3.3.5 Specular Reflection

Once eyelid occlusions are estimated, occlusions resulting from specular reflection are evaluated on the remaining iris portion unaffected by the eyelids. This factor is estimated by traditional hard thresholding. Based on evaluation of ICE 1.0 and WVU data sets, a threshold of 240 experimentally gave good results.

Fig.3.12 contains a normalized iris image and the corresponding specular reflection mask generated by traditional hard thresholding.

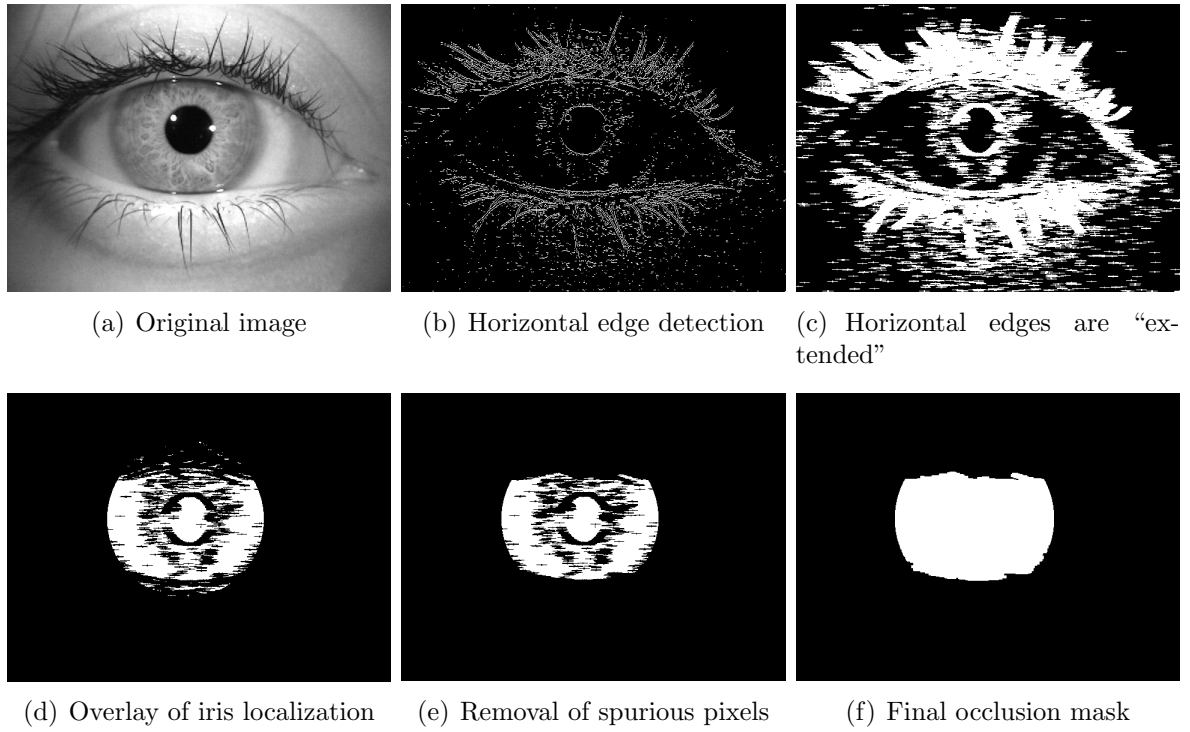


Figure 3.11: Occlusion mask generation.

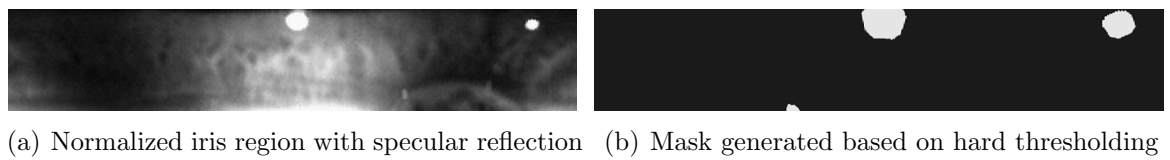


Figure 3.12: Illustration of specular reflection mask generation.

3.3.6 Lighting Variation

After estimating occlusions from eyelids and specular reflection, the remaining unoccluded iris portion is split into twelve regions. A mean is calculated in each region, X_i .

$$\begin{aligned} \text{Lighting} &= \frac{\sum (X_i - \mu_R)^2}{N}, \\ \text{where } \left\{ \begin{array}{l} \mu_R &= \frac{1}{N} \sum X_i \\ N &= \# \text{ of regions} \end{array} \right. \end{aligned} \quad (3.5)$$

μ_R is the mean of all regions $X_{i...N}$. Finally the variance of the means is used as our estimate for lighting variation, as in (3.5).

3.3.7 Pixel Counts

Our measure for pixel counts is calculated as the ratio of iris pixels to the combined number of iris and occluded (resulting from eyelids, eyelashes, specular reflection) pixels within the iris region:

$$\text{Pixel Counts} = \frac{X_{estimated}}{X_{estimated} + X_{occluded}}, \quad (3.6)$$

where $X_{occluded}$ represents the number of pixels occluded from eyelids and specular. $X_{estimated}$ represents the number of estimated iris pixels. Although resolution is an important factor as illustrated in the 3.2, we do not incorporate it into this specific measure as publicly available data sets have acceptable resolutions above the ISO standard (200 pixel iris diameter).

3.4 Combining Quality Factors

Typically, vendors of biometric systems as well as the organizations which use iris recognition on a large scale, desire to have a unified quality score as a measure of overall suitability for authenticating an individual. Further, recent “quality-enhanced” multi-modal biometric fusion algorithms [70, 75, 106] tend to use a single quality estimate for each biometric modality included in the fusion framework. To generate an overall quality of iris images based on the estimated individual factors, we adopt an approach based on Dempster-Shafer (DS) theory [96]. This approach was proposed as a solution to a number of problems in the fields of artificial intelligence, software engineering, and pattern classification. We utilize DS theory because of its low complexity (compared to say Bayesian theory), explicit treatment of statistical dependence between dimensions (compared to, say, Fuzzy logic approaches) and its demonstration of good performance in many applications such as robotics [74], software fault prediction [41] and biometrics [107]. We evaluated other approaches to fuse iris quality estimates (namely, Bayesian networks and Fuzzy logic), but due to limited space, we discuss only the DS approach. It is worth mentioning that we found the DS to be the most straightforward, in terms of low performance overhead and an intuitive description, and that in our experiments other fusion algorithms did not offer any performance advantages.

DS theory relies on concepts of beliefs, propositions, and evidence. The belief for propositions (similar to events in Bayesian theory) start at 0, with uncertainty equal to 1. Based on incoming evidence, belief assignments are updated. This results in decreasing the uncertainty. In DS theory, belief models are built on a finite boolean algebra of mutually exclusive propositions known as the frame of discernment, denote it by Θ . The belief in a proposition

$Bel(A)$ is a measure of certainty that evidence A is true. Shafer [96] gives the following expressions for assigning and measuring beliefs.

If Θ is a frame of discernment, then a function $m : 2^\Theta \rightarrow [0, 1]$ is called a basic probability assignment when:

1. $m(\emptyset) = 0$
2. $\sum_{A \subseteq \Theta} m(A) = 1.$

To measure the belief of a proposition A , one must add up the belief in all subsets B belonging to A :

$$Bel(A) = \sum_{B \subseteq A} m(B). \quad (3.7)$$

Dempster's rule of combination is used to combine beliefs over the same frame of discernment that are from distinct sources of evidence. This is measured by computing the orthogonal sum of all belief functions m which results in a new belief function based on the combined evidence:

$$m(C) = \frac{\sum_{A_i \cap B_j = C} m_1(A_i)m_2(B_j)}{1 - \sum_{A_i \cap B_j = \emptyset} m_1(A_i)m_2(B_j)}, \quad (3.8)$$

where $m(C)$ is the new belief resulting from the combination of the beliefs for evidence's A and B . Note that Dempster's rule assumes that the evidence are independent. The problem with this assumption in our application lies in the fact that we do not have a good understanding of the dependencies between the quality factors and to assume independence between them is unreasonable (since our evidence is from the same source). In light of this, Murphy [73, 74] modified Dempster's rule such that it is suitable to use information from the

same source as seen in equations (3.9) and (3.10):

$$m(C) = \frac{\sum_{A_i \cap B_j = C} f(m_1(A_i)m_2(B_j))}{1 - \sum_{A_i \cap B_j = \emptyset} f(m_1(A_i)m_2(B_j))}, \quad (3.9)$$

where $f(\cdot)$ is given by:

$$f(m_1(A_i)m_2(B_j)) = [m_1(A_i)m_2(B_j)]^n, n \in [0, 1], \quad (3.10)$$

in Murphy's rule. Murphy characterizes n as a method to weight evidence. Choosing $n > 0.5$ gives more weight when combining new evidence, while choosing $n < 0.5$ gives less weight when combining new evidence [74]. Other proponents of Murphy's rule characterize n as governing correlation between evidence [71]. It is explained in [71], that choosing $n > 0.5$ assumes more independence between the evidence while choosing $n < 0.5$ assumes correlation. In light of both views, choosing $n = 0.5$ is considered neutral and equal weight is applied to all evidence during integration.

3.4.1 Quality Score Normalization

Prior to fusing the factors, they need to be normalized between $[0, 1]$ [71, 73]. Occlusion, specular, and pixel counts are ratios which are already between the desired score range. Defocus and motion blur are normalized based on a modified form of min-max normalization:

$$Q_{new} = \frac{Q_{old} - Q_{min}}{Q_{max} - Q_{min}}. \quad (3.11)$$

With respect to defocus, Q_{old} represents the raw power scores obtained from the bandpass kernel. In the case of motion, Q_{old} is the raw power score corresponding to the localized peak which is extracted from directional filter that is perpendicular to the estimated angle. Power constants Q_{max} and Q_{min} represent the maximum and minimum values for which the quality scores are normalized between. These values were learned based on a subset of images from ICE 1.0. They were not changed when evaluating other data sets. Recall that lighting variation is just the variance of the means extracted from un-occluded iris regions. However, inaccurate occlusion estimation can cause outliers in the estimate resulting in a score greater than one. In this case, the score is simply set to one.

3.4.2 Dempster Shafer Theory Applied to Quality Assessment

We adopt a frame of discernment containing two propositions which represent opposite beliefs:

- A - Image quality is bad (our belief that quality is bad).
- B - Image quality is good (our belief that quality is good).

The normalized values for each quality factor are assigned as beliefs to proposition A. Since these propositions represent opposite beliefs, the assigned belief to B is essentially the complement of the assigned belief to A . We adopt Murphy's rule to combine beliefs with parameter $n = 0.5$ for all evidences. A generalized expression for combining beliefs from k quality

factors m_1 to m_k is given by:

$$\hat{m}_i(A) = \frac{(m_{i-1}(A) \cdot m_i(A))^n}{((m_{i-1}(A) \cdot m_i(A))^n + (m_{i-1}(B) \cdot m_i(B))^n)}, \quad (3.12)$$

$$i = 2, \dots, k,$$

where $m_i(B) = 1 - m_i(A)$, since our propositions are complements of each other. Murphy has shown that different orderings result in different values for combined beliefs [74]. Since we have seven quality factors, that will result in $7!$ combinations. Our goal is to attain the orderings that result in the minimum and maximum values. These values are important because we can consider them as the worst case and best case quality for a specific image. Mladenovski [71] has proved that a maximum value can be attained by first sorting the beliefs in ascending order with $n = 0.5$. Similarly, if sorted in descending order a minimum value can be obtained. The following section illustrates application of the fusion rule on samples from WVU, CASIA 3.0 and ICE 1.0 data sets.

3.4.3 Evidence Fusion Examples Based on Murphy's Rule

The sample iris images in Fig. 3.13 are from WVU, ICE 1.0, CASIA 3.0 data sets. Image (a) represents a good quality image from WVU and (c) represents a good quality image from ICE 1.0 (based on visual evaluation). Images (e) and (f) represent good and poor quality images from CASIA 3.0 respectively. Images (b) and (d) represent degraded quality images which are effected by motion (b) and occlusion, defocus, and lighting (d). Image (f) scores low because the iris is flat with almost no visible texture. The estimated direction of motion blur for Fig. 3.13 (b) is 85° counter-clockwise with respect to the pupil center being at 0° .

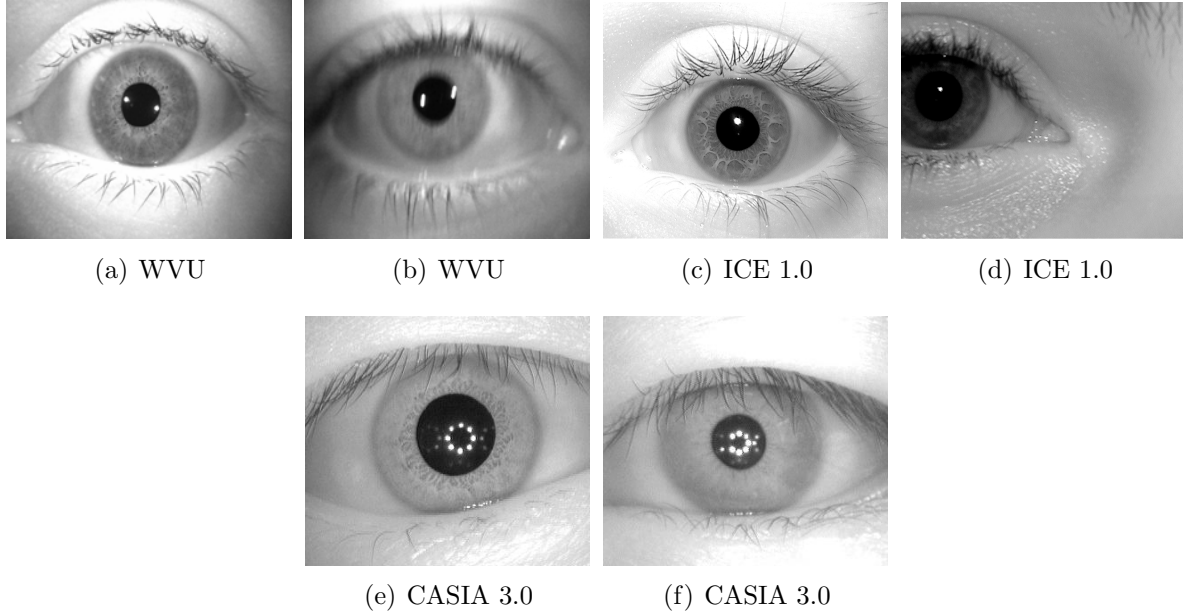


Figure 3.13: Sample Images from WVU, ICE 1.0, and CASIA 3.0 data sets

Table 3.2 lists the estimated factors (factors are normalized to take values between 0 and 1, with 1 implying heavy degradation) for these images and the combined quality for them. The quality column (last column in the table) represents the lower value (minimum value attained from fusion of all factors which we loosely consider as the worst case quality) on image quality. We use this value as our global quality score.

Image	Factor	Defocus	Motion	Occlusion	Specular	Lighting	Pixel Count	Quality
(a)		0.01	0.01	0.01	0.00	0.01	0.01	0.97
(b)		0.27	0.66	0.04	0.00	0.21	0.05	0.63
(c)		0.00	0.00	0.08	0.00	0.03	0.08	0.95
(d)		0.83	0.17	0.55	0.00	0.99	0.54	0.08
(e)		0.09	0.10	0.06	0.03	0.01	0.08	0.91
(f)		0.90	0.13	0.20	0.03	0.06	0.2	0.19

Table 3.2: Estimated Factors for images in Fig. 3.13

3.5 Numerical Results

In this work we performed analysis using three iris databases, CASIA 3.0 [1], WVU, and ICE 1.0 [61]. For CASIA 3.0 we used a subset of images called “Interval”. This data consisted of 2655 images captured at a resolution of 280x320. The WVU data set consists of 2413 images taken at a resolution of 640x480. ICE 1.0 consists of 2953 images captured at a resolution of 640x480. Images that failed segmentation completely, i.e. no localization for pupil and iris are removed (if this occurs, quality is set to 0). On the other hand, images that segmented but have incorrect localization (based on visual evaluation) are kept for the following experimentation. Table 3.3 lists the segmentation results for the data sets used during experimentation. Fig. 3.14 illustrates some images with correct and incorrect localization.

Dataset	Image Count	Complete Failure	Incorrect Localization	Final Count
CASIA 3.0	2655	8	17	2647
WVU	2413	0	50	2413
ICE 1.0	2953	1	30	2952

Table 3.3: Segmentation statistics for all three data sets.

3.5.1 Quality Factor Distributions

This section contains histograms of the estimated quality factors for CASIA 3.0, WVU , and ICE 1.0 data sets. The x-axis for all plots represents the strength of the estimated factor normalized to $[0,1]$. Statistics for these distributions is illustrated in Table 3.4. Fig. 3.15 presents histogram plots of defocus and motion blur estimates for the data sets. All three data sets have a wide range of defocus scores, with ICE 1.0 being affected more than the

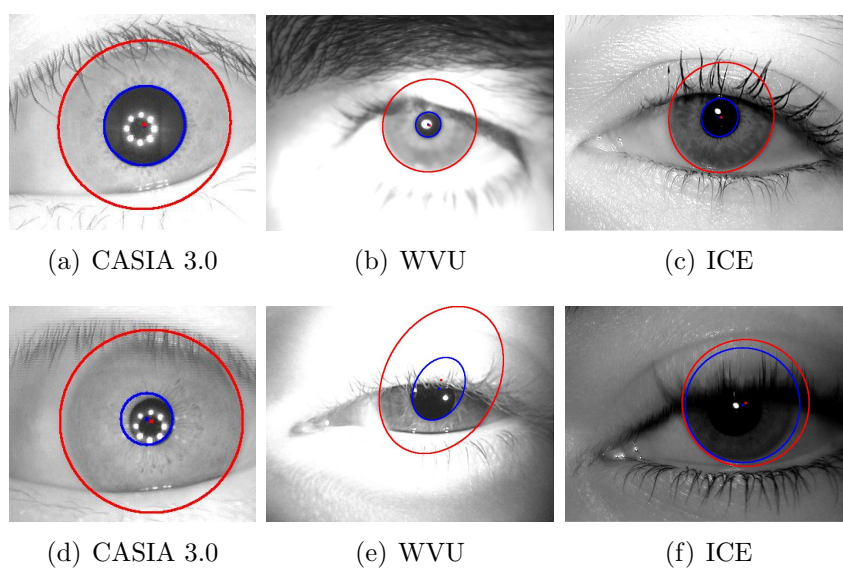


Figure 3.14: We classify images as incorrect if pupil and iris segmentation boundaries do not fall completely along their respective boundaries. (a), (b), (c) represent correct localization results from CASIA 3.0, WVU, and ICE 1.0 respectively. (d), (e), (f) represent incorrect localization from CASIA 3.0, WVU, and ICE 1.0 respectively.

other data sets.

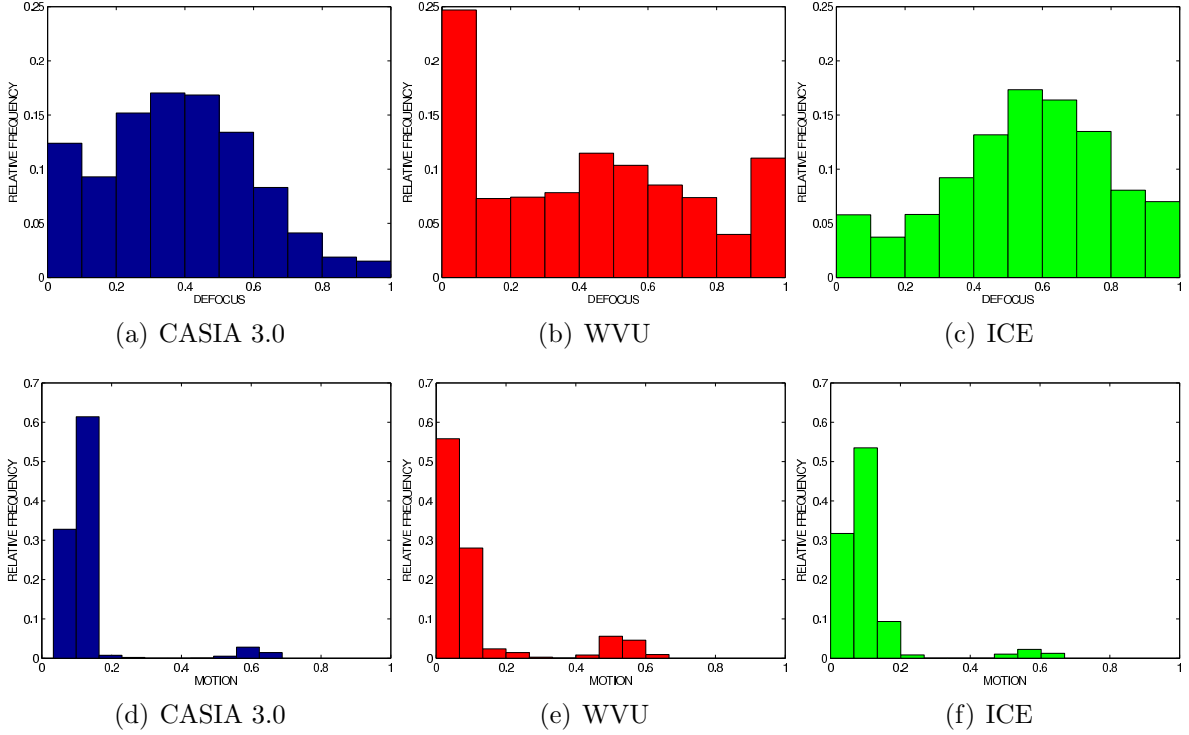


Figure 3.15: Defocus and motion frequencies for CASIA 3.0, WVU, and ICE

The histograms for motion blur indicate that none of the data sets under study are significantly effected by this factor. Fig. 3.16 shows histogram plots of occlusion and lighting variation estimates. For all three data sets there are moderate amounts of occlusion. The CASIA 3.0 data set contains the highest occlusion scores, with a mean occlusion estimate of 0.1999 and a variance of 0.0164. This can be verified when visually inspecting the data set. In terms of lighting variation, WVU has the most lighting variation with a mean lighting score of 0.2140 and a variance of 0.0502. CASIA 3.0 has the least amount of lighting variation with a mean of 0.0451 and variance of 0.0016. Fig. 3.17 characterizes the distributions for

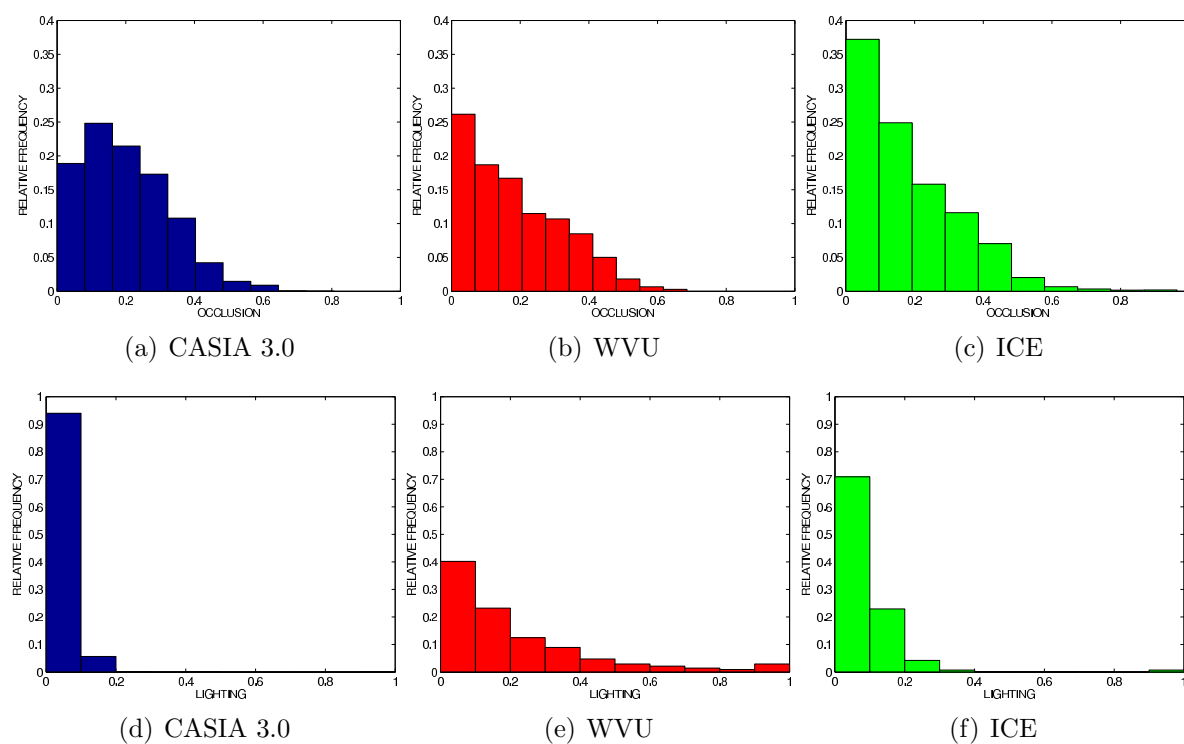


Figure 3.16: Occlusion and lighting frequencies for CASIA 3.0, WVU, and ICE

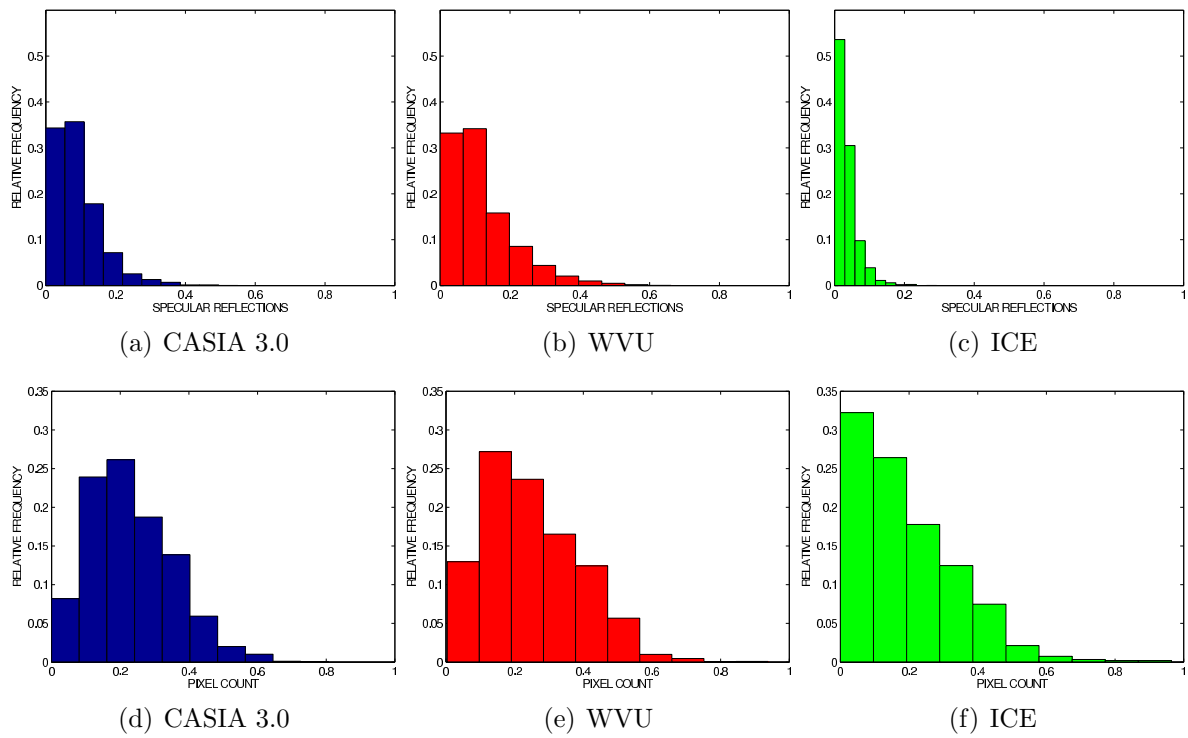


Figure 3.17: Specular and pixel count frequencies for CASIA 3.0, WVU, and ICE

specular reflection and pixel count estimates. WVU has the highest frequency of specular reflection in the iris texture, with a mean of 0.1192 and variance of 0.0085. ICE 1.0 on the other hand, has the lowest frequency of specular reflection in the iris texture, with a mean of 0.0347 and a variance of 0.0093. Similar to occlusion, all data sets suffer from low pixel count scores. WVU has the highest frequency of pixel count while ICE 1.0 has the lowest.

Fig. 3.18 illustrates the distribution of quality (minimum value attained from Murphy's rule with $n = 0.5$) for (a) CASIA 3.0, (b) WVU, and (c) ICE 1.0 databases. CASIA 3.0 has the highest quality data followed by ICE 1.0, then WVU. WVU suffers from many factors such as defocus, occlusion, and lighting, which accounts for low scoring data. If we were to rank these data sets in terms of quality based on these plots, CASIA 3.0 scores the highest, followed by ICE 1.0, and then WVU.

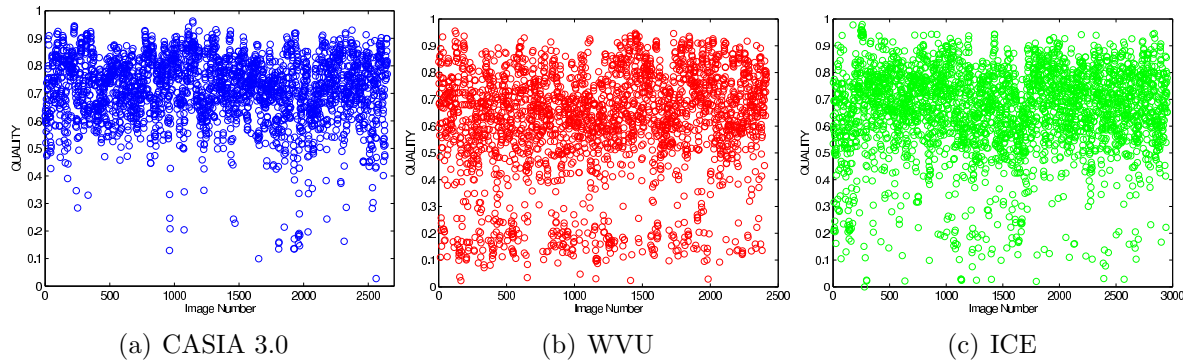


Figure 3.18: Overall quality scatter plots for CASIA 3.0, WVU, ICE 1.0 data sets

Table 3.4 summarize mean and variance scores for CASIA 3.0, WVU, and ICE 1.0 data sets (mean followed by variance).

Dataset	Factor	Defocus		Motion		Occlusion		Specular		Lighting		Pixel Count	
CASIA 3.0		0.3788	0.0491	0.1302	0.0121	0.1999	0.0164	0.0902	0.0047	0.0451	0.0016	0.2319	0.0147
WVU		0.4219	0.1070	0.1195	0.0249	0.1841	0.0199	0.1192	0.0085	0.2140	0.0502	0.2499	0.0195
ICE 1.0		0.5528	0.0594	0.1029	0.0126	0.1768	0.0225	0.0347	0.0093	0.0879	0.0108	0.1922	0.0214

Table 3.4: CASIA 3.0, WVU, and ICE 1.0 Mean and Variance Quality Factor Scores

3.5.2 CASIA 3.0, WVU, and ICE 1.0 Performance

Performance of biometric systems is typically characterized by error rates such as false accept rate (FAR), genuine accept rate (GAR), and equal error rate (EER). To measure the separation of the genuine and imposter distributions we employ the d' metric [8]. To evaluate the performance of our quality metric, we divide each data set into three quality subsets (low, medium, high) such that each interval has same number of samples. This is done so that there is no bias between the intervals which may result when the intervals do not contain the same amount of samples. The interval ranges for each data set is presented in table 3.5.

Dataset	Quality Interval	Low	Medium	High
CASIA v3.0		0 - 0.6779	0.6779 - 0.7769	0.7769 - 1
WVU		0 - 0.5759	0.5759 - 0.7144	0.7144 - 1
ICE 1.0		0 - 0.6162	0.6161 - 0.7429	0.7429 - 1

Table 3.5: Quality interval ranges for each data set.

Fig. 3.19 contains three ROC plots characterizing the performance of each data set for the three different quality intervals. With all three data sets, the higher quality data scores better than the lower quality data.

Tables 3.6 and 3.7 verify the plots by illustrating the performance statistics for the three different data sets. Note that CASIA 3.0 data is of higher quality than the other two data

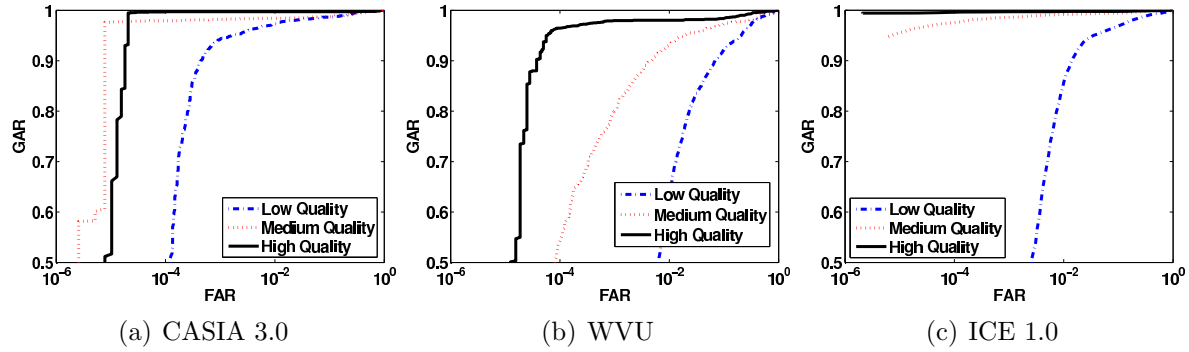


Figure 3.19: Verification performance for all three data sets.

sets which is corroborated with the lowest EER and highest d' . The ICE 1.0 data scores higher than WVU but lower than CASIA 3.0 which is also validated by its EER and d' .

Interval	EER %	d'	Mean Quality	Image Count
Low	0.0199	4.5604	0.5823	882
Medium	0.0157	5.7526	0.7286	883
High	0.0028	8.1291	0.8349	882

Interval	EER %	d'	Mean Quality	Image Count
Low	0.0870	2.4766	0.3765	804
Medium	0.0407	3.1934	0.6451	805
High	0.0189	4.9789	0.8059	804

Table 3.6: Summary of performance statistics using Log-Gabor encoding algorithm for the CASIA 3.0 data set and WVU data set.

Interval	EER %	d'	Mean Quality	Image Count
Low	0.0446	3.4886	0.4789	984
Medium	0.0081	5.3096	0.6830	984
High	0.0029	6.3622	0.8126	984

Table 3.7: Summary of performance statistics using Log-Gabor encoding algorithm for the ICE 1.0 data set.

3.6 Performance Comparison

In order to further demonstrate the effectiveness of our proposed quality metric we compare it to our implementation of two global iris quality metrics, namely Daugman's focus metric [28]

and Kang et. al.'s focus metric [53]. Additionally, we also tested the metric employed in [18], however we could not reproduce the results from [18] hence we omit the experimentation done with this metric.

Each metric is tested on CASIA v3.0 and WVU non-ideal data. The images in each data set are partitioned into two equal subsets, high quality and low quality. A set of images consisting of the entire data set is scored independent of quality which is used as a baseline. Intuitively, metrics capable of performance prediction should score better than the baseline when evaluating their high quality intervals since this interval pertains to only high quality data from the baseline set of images. Alternatively, the low quality intervals should score below the baseline.

3.6.1 Experimental Results

Fig. 3.20 illustrates the performance prediction results for the metrics under study using the CASIA v3.0 data. Fig. 3.20 (a) and (b) are ROC plots of the high quality and low quality intervals for all metrics respectively. It is quite clear that the proposed metric is

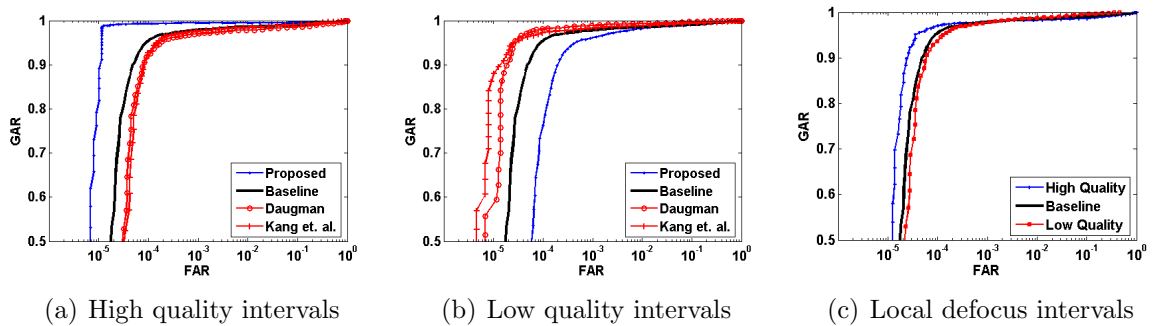


Figure 3.20: CASIA V3.0 Performance prediction comparison

capable of performance prediction as the high quality interval scores above the baseline and the low quality interval scores below the baseline. On the other hand, the performance of the high and low quality intervals for the global metrics are transposed (i.e. high quality interval scores below the baseline and low quality interval scores above the baseline). This interesting phenomenon might suggest that using defocus alone as a quality metric is not enough to reliably predict performance. Alternatively, this might suggest that the global focus metrics are being influenced by heterogenous factors in the iris image not part of the iris such as eyelashes, eyebrows, or even the sclera. To further explore these possibilities consider Fig. 3.20 (c) which is an ROC plot of high and low quality intervals using the proposed defocus metric. Recall that the proposed defocus metric is simply Daugman's metric but applied locally to iris texture rather than globally. The high quality interval scores above the baseline and the low quality interval scores below the baseline which implies that the proposed local defocus metric by itself is indicative of performance prediction. Now consider Fig. 3.21 (a)

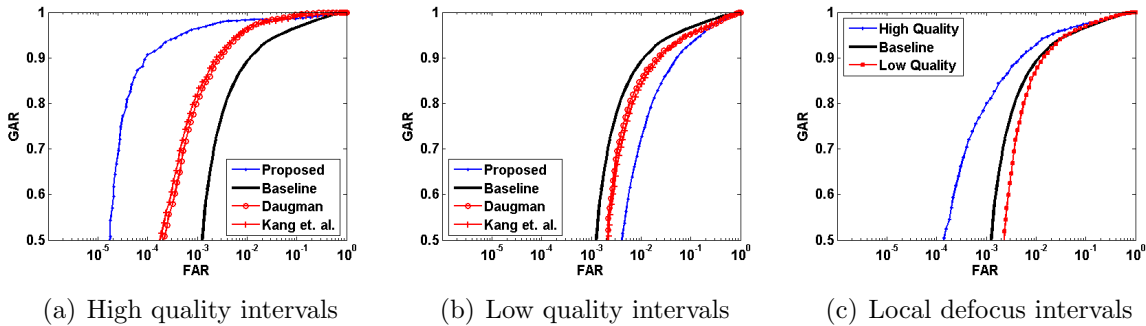


Figure 3.21: WVU Performance prediction comparison

and (b) which are ROC plots of high and low quality intervals for the WVU non-ideal data set. The global metrics and the proposed metric are indicative of performance prediction for both intervals. Furthermore, consider Fig. 3.21 (c) which is an ROC plot of prediction using

the proposed defocus metric. These plots suggest that the global metrics are able to provide performance prediction. However, this conflicts with the results from the CASIA data. We conjecture that the conflict arises because CASIA has more occlusion than WVU non-ideal (see table 3.4). As a result, the global focus metrics are influenced more from heterogenous non-iris regions rather than the iris region.

3.7 Chapter Summary

In this chapter quality factors for the iris biometric were identified and their individual influence on performance was studied for the Log-Gabor based encoding technique. We concluded that for all quality factors identified, they negatively influence recognition performance with the main degradation observed in the genuine scores as the degree of degradation increases.

Estimation procedures for defocus, motion blur, occlusion, specular reflection, lighting, pixel counts were then carried out. Estimated quality scores are then fused by use of Dempster-Shafer theory in order to attain a combined metric for image quality. We tested the proposed quality algorithm on CASIA 3.0, WVU, and ICE 1.0 data sets. We noticed that CASIA 3.0 contains higher quality data than both ICE 1.0 and WVU. We empirically observed that our quality metric can predict performance for all three data sets, with improvements in both EER and d' as quality increases. This result indicates that our fused iris quality measure is suitable as an informal measure of dependability of the matching decision. As such, the measure can be included in multi-modal fusion algorithms which utilize quality measures from each modality. Furthermore, we compared our quality metric to global iris quality metrics and concluded that although they are capable of predicting performance,

local metrics provide better prediction despite their obvious limitation of requiring segmentation. Our metric also has the advantage of providing descriptive feedback about the quality of the image (i.e. we can describe how the image is degraded and to what extent). This is useful when deciding another capture is required and can provide the operator with feedback regarding how the settings should be adjusted to capture a better quality image.

The main limitation of our approach is the requirement of segmentation. Failed localization/segmentation will result in inaccurate quality scores. However, this would have negative consequence only if the matching algorithm applied to the same iris image performs segmentation successfully. Therefore, as long as the segmentation algorithm used for quality evaluation is as sophisticated as the one used in quality evaluation, it is unlikely that quality scores will be misleading [128]. Nevertheless, the need to deploy segmentation within the quality assessment algorithm makes this approach unsuitable for real-time applications in which a quality factor would be used for the selection of the “best” frame from a sequence (e.g., streaming video).

Aside from perfecting the estimation techniques for the described quality factors, experimenting with the new quality scores that incorporate correlation may prove useful. Further, the proposed framework is open for the inclusion of new iris quality factors which will undoubtedly emerge through further research [45] or through further relaxation of acquisition constraints (e.g. distance, motion, non-uniform lighting) [84]. To support such future development, we will maintain and expand our data sets, keep the data as well as the described algorithms publicly available and, therefore, encourage repeatability of our experimental results.

Chapter 4

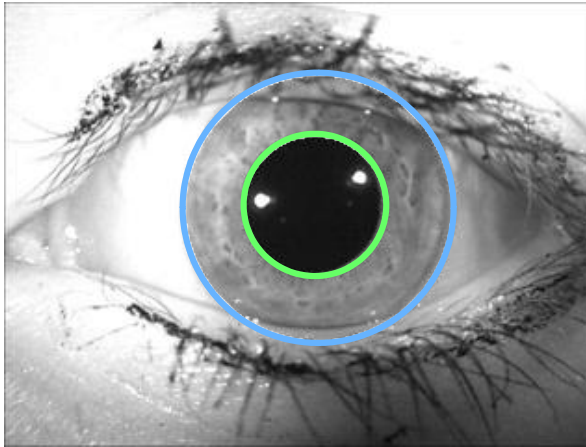
Segmentation Evaluation

4.1 Introduction

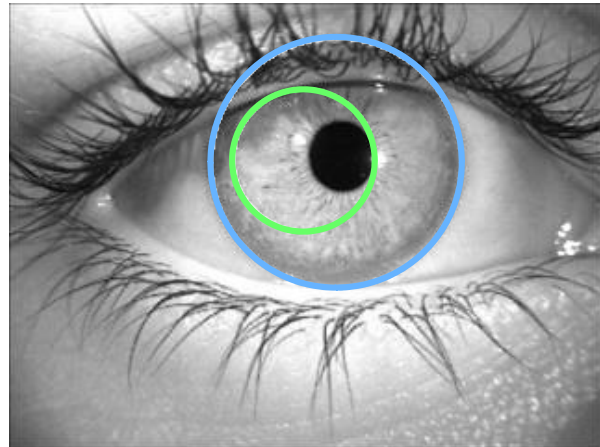
The performance of iris recognition systems is driven in part by application scenario requirements. Standoff distance, subject cooperation, underlying optics, and illumination are just a few examples of factors associated with these requirements. These factors subsequently dictate the nature of images an iris recognition system will deal with. At the image level, iris segmentation is arguably one of the most important factors driving recognition performance. That is, if the iris regions are successfully localized for pairs of images to be matched, the correct classification will be almost always be made. Iris image segmentation typically consists of two problems. First, one must define the boundary between the pupil, the black region in the center of the eye and the iris, the textured region surrounding the pupil as shown by the inner green ring in Figure 4.1(a). Second, one must define the boundary between the iris and the sclera, or the lighter region surrounding the iris as shown by the outer blue ring in Figure 4.1(a). Many methods exist for detecting these boundaries, a survey of segmentation

algorithms is provided in Chapter 5.3.

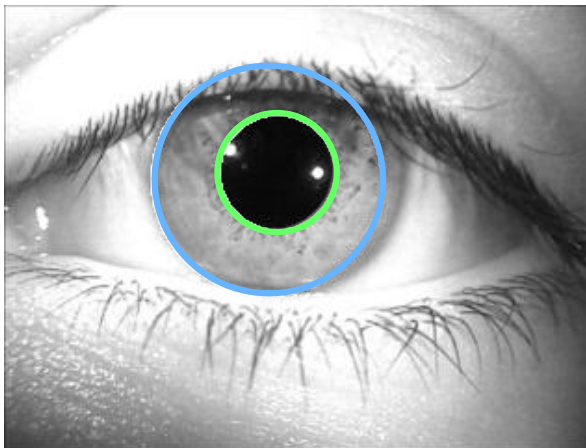
Whether due to limitations of algorithms or poor image quality, failed segmentation



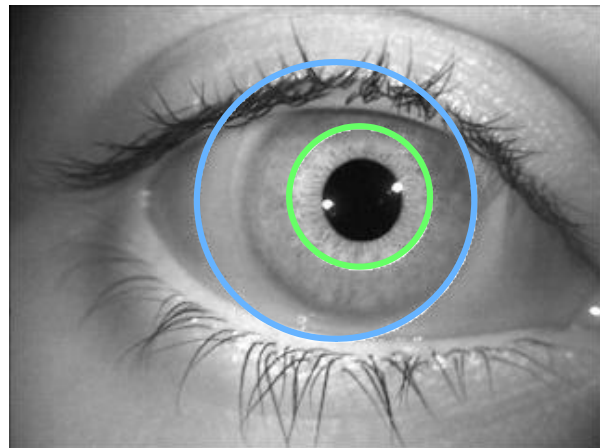
(a) Correctly segmented image.



(b) Failed pupil segmentation.



(c) Failed iris segmentation.



(d) Failed pupil and iris segmentation

Figure 4.1: Four types of segmentation results (a) Correctly segmented image (b) Failed pupil segmentation (c) Failed iris segmentation (d) Failed pupil and iris segmentation.

often accounts for misclassification errors in iris recognition systems. As a result, the ability to automatically evaluate whether the segmentation block of an iris recognition system has

succeeded or failed is of paramount importance when attempting to predict the outcome of matching. Whether a binary success / failure flag or measures with higher granularity, currently existing algorithms do not explicitly evaluate the segmentation result. As a result, without human inspection, the success of segmentation blocks is largely unknown in most iris recognition systems. Having a tool which provides such information is useful in an operational sense in that it can serve as an indicator to reacquire a better image if feasible. Otherwise, when reacquisition is not an option, such a measure could serve to flag entrance into a computationally more expensive automatic segmentation block (e.g., segmentation rectification).

In this chapter, we provide a look into the feasibility of automatically evaluating the segmentation block of iris segmentation algorithms. Our segmentation evaluation methodology is a combination of metrics that utilize geometric, statistical intensity, and edge information. Besides looking at the precision of the pupil and iris segmentation boundaries independently, we also provide the ability to arrive at a global binary segmentation evaluation result by way of a machine learning approach. We test the accuracy of the approach on three databases using three different segmentation algorithms. Additionally, to demonstrate one application of the tool, we investigate the effect that the varying success of segmentation has on iris matching performance.

4.2 How do we evaluate segmentation?

In general, image segmentation is the process of partitioning an image into salient regions or segments such that the overall representation is easier to understand and analyze. Con-

siderable research in the field has yielded segmentation methodologies for a wide range of applications. However, the problem of evaluating which algorithm produces more accurate segmentations than another is still an open research problem. Traditional segmentation evaluation strategies fall into three categories: subjective, supervised, and unsupervised. Subjective evaluation strategies rely on a human to visually compare and interpret image segmentation results for separate segmentation algorithms. Thus, analysis is a time consuming process and limited to a small number of actual comparisons. Supervised evaluation methodologies, where segmentations are compared to manually annotated ground truth images, still require user assistance; Therefore analysis is still limited by human interpretation. Unsupervised evaluation strategies do not require a human in the loop, thus enabling an objective comparisons for different segmentation algorithms and different parameterizations of the same algorithm [104,121]. In this chapter we focus on developing and testing a supervised segmentation evaluation machine learning algorithm.

4.3 Approach

Our approach to automatic iris segmentation evaluation utilizes geometric, statistical intensity, and edge information to evaluate the estimated pupil and iris boundaries. Geometric information is primarily used to discern gross segmentation failures such as those illustrated in Figure 4.1(b) and 4.1(d). Statistical intensity information is primarily used to evaluate the pupil boundary while edge information is used primarily for evaluating the precision of the iris boundary. Finally, all scores are utilized as features to build an ensemble through machine learning, to evaluate the overall decision of whether the segmentation process has

succeeded or failed. The following text describe these metrics in more detail.

4.3.1 Geometric Measures

Often, the pupil and iris boundaries are not necessarily concentric but the distance between their centers is typically small [27] (with the exception of extreme off-angle irides). Observations from grossly inaccurate segmentation boundaries illustrate that the distance between the pupil and iris center is unusually large. Based on this observation, we make use of concentricity of the pupil and iris boundaries. A similar observation can be made from segmentation models that represent the shape of each boundary as an ellipse. Furthermore, the eccentricity of the ellipse can be utilized to characterize gross segmentation failures as suggested in [49]. The following expressions can be used as measures which quantify gross segmentation failures for circular (4.1) and elliptical (4.2) based models:

$$G_c = \sqrt{(p_x - i_x)^2 + (p_y - i_y)^2}, \quad (4.1)$$

$$G_e = \sqrt{(p_x - i_x)^2 + (p_y - i_y)^2} + \arccos\left(\frac{b_i}{a_i}\right) * 100 + \arccos\left(\frac{b_p}{a_p}\right) * 100, \quad (4.2)$$

where (p_x, p_y) are the pupil center coordinates, (i_x, i_y) are the iris center coordinates, b_i and a_i are the semi-minor and semi-major axes for iris ellipse and b_p and a_p are the semi-minor and semi-major axes for pupil ellipse.

It is easy to envision how the aforementioned geometric measures apply straightforwardly to gross inaccuracies in the segmentation boundaries such as those illustrated in Figure 4.1(b) and 4.1(d). However, consider the image in Figure 4.1(c), which illustrates an incorrect iris

boundary. Note that in this case, the offset between centers of both boundaries is relatively small compared to Figure 4.1(b) and 4.1(d). In this instance, the prescribed geometric measures may not provide enough information to differentiate as an incorrect segmentation. This limitation can be mitigated by including geometric information extracted from the annulus generated from both pupil and iris boundaries. For instance, observe the annulus in the upper 12 to 3 o'clock position of Figure 4.1(c). Note that this region has radial distances much smaller than the annulus region in the 6 to 9 o'clock position. This information can be exploited to differentiate the correctness of the boundaries. More specifically, radial distances are measured from the annular region at “symmetric” sampling angles. That is, distances are measured on the lower hemisphere of the annulus at angles, $L_\theta \in [0, 180]$, and on the upper hemisphere of the annulus at its corresponding “symmetric” angle, $U_\theta = L_\theta + 180$. The estimated distances are then transformed to ratios by dividing the smaller distance by the larger for each corresponding sampling angle. Mathematically, this is described as follows.

$$\overline{G}_\theta = \begin{cases} \frac{U_\theta}{L_\theta}, & \text{if } L_\theta > U_\theta \\ \frac{L_\theta}{U_\theta}, & \text{otherwise.} \end{cases} \quad (4.3)$$

Here, \overline{G}_θ , is a vector of annular distance ratios characterizing the shape of the annular region created between the pupil and iris boundary. Annular distance ratios approaching 1 indicate higher concentricity or better segmentation while the opposite is true for ratios approaching 0. This process is illustrated in Figure 4.2, where the top row illustrates estimated segmentation boundaries that are grossly inaccurate while the bottom has only a small offset of the pupil boundary. In the third column, annular ratios are provided for both images.

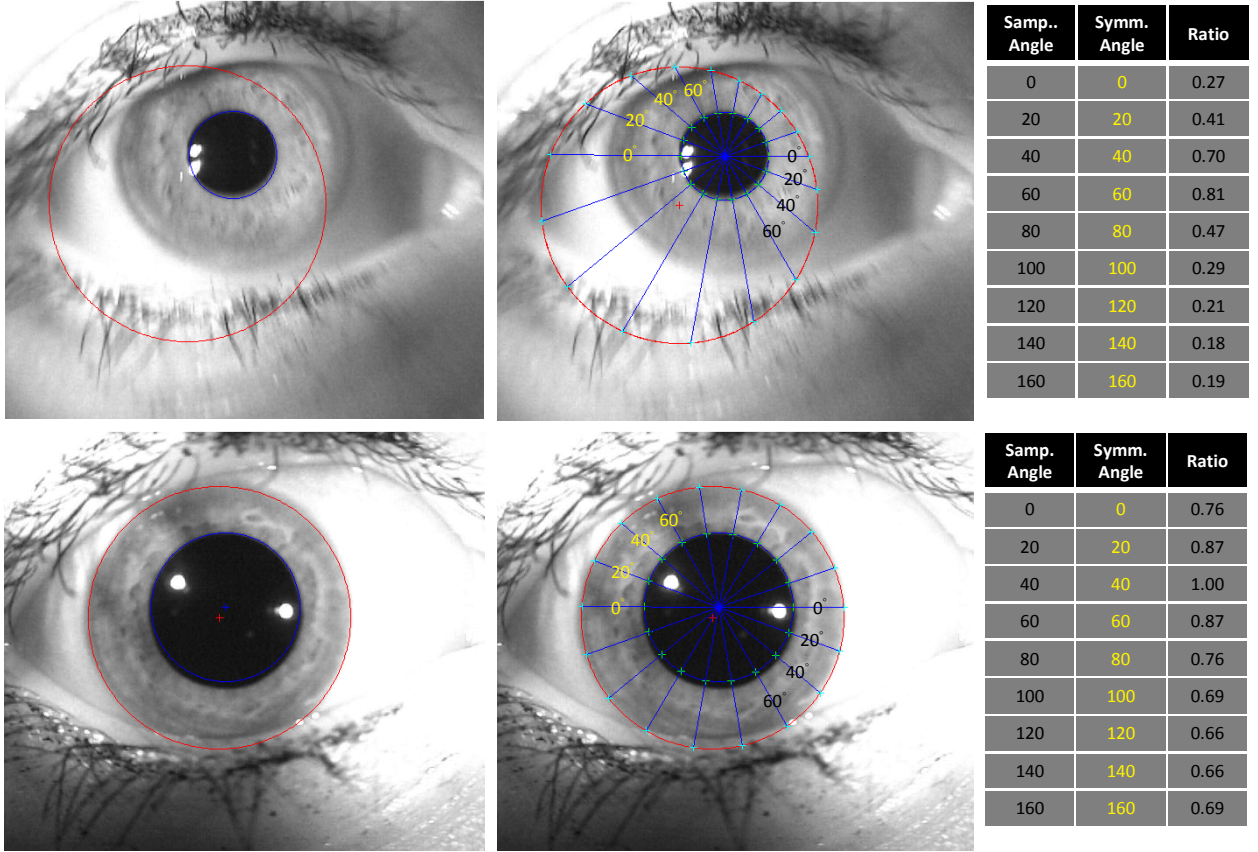


Figure 4.2: Illustration of annular distance ratio. The image in the top row is grossly inaccurate with respect to the iris boundary. The corresponding distance between the boundary centers is over 40 pixels while the boundary centers in the bottom row image are offset by only 13 pixels.

4.3.2 Statistical Intensity Measure

Beyond geometric information, we also make use of intensity information pertaining to the pupillary region. Several researchers [27, 60, 130] have empirically observed that the pupil region, when illuminated with near infrared, elicits stronger characteristics, such as increased contrast over the iris/sclera region. Thus, segmentation algorithms typically proceed by attempting to segment the iris/pupil region first because it is relatively more stable (e.g.,

more tolerance to non-ideal conditions) in comparison to the iris/sclera region. Thus, we make use of the intensity information from the pupil region to discern the correctness of the pupil boundary. More specifically, given a pupil segmentation boundary, we are interested in whether the pixels, \bar{x} , that fall within the pupillary boundary are indeed pupil pixels. By fitting probabilistic models for \bar{x} , we can formulate a likelihood ratio test, $\Lambda(\bar{x})$, to decide between pupil and non-pupil pixels. In other words,

$$\Lambda(\bar{x}) = \frac{P(\bar{x}|H_1)}{P(\bar{x}|H_0)} \geq \eta, \quad (4.4)$$

where $H_1 : \bar{x}$ corresponds to pupil pixel and $H_0 : \bar{x}$ corresponds to a non-pupil pixel. To do so, we need to choose models for $P(\bar{x}|H_1)$ and $P(\bar{x}|H_0)$. We assume that the pupil area of an iris image is a relatively stable, flat homogenous region of dark intensities (with respect to the iris), with discontinuities resulting only in the presence of eyelashes / eyelids or specular reflection. Additionally, we observe from the spatial intensity histograms of correctly segmented pupil regions, under ideal lighting conditions, that most of the intensity frequencies fall close to zero with a small step-ladder effect trailing off towards higher intensities as a result of specular reflection / eyelashes. On the other hand, under ideal lighting conditions, we assume the iris area is much more heterogenous as compared to the pupil area, particularly around the collarette region, while the iris region closer to the sclera is much flatter. The spatial histograms of correctly segmented iris regions (without occlusion masks) show that the frequency spread is much wider than that of the pupil and curiously bell-shaped. Given these observations and based on empirical evaluation, we fit a Gamma distribution for $P(\bar{x}|H_1) \sim \Gamma(k, \Theta)$ which is characterized by two parameters, shape k and scale Θ . We

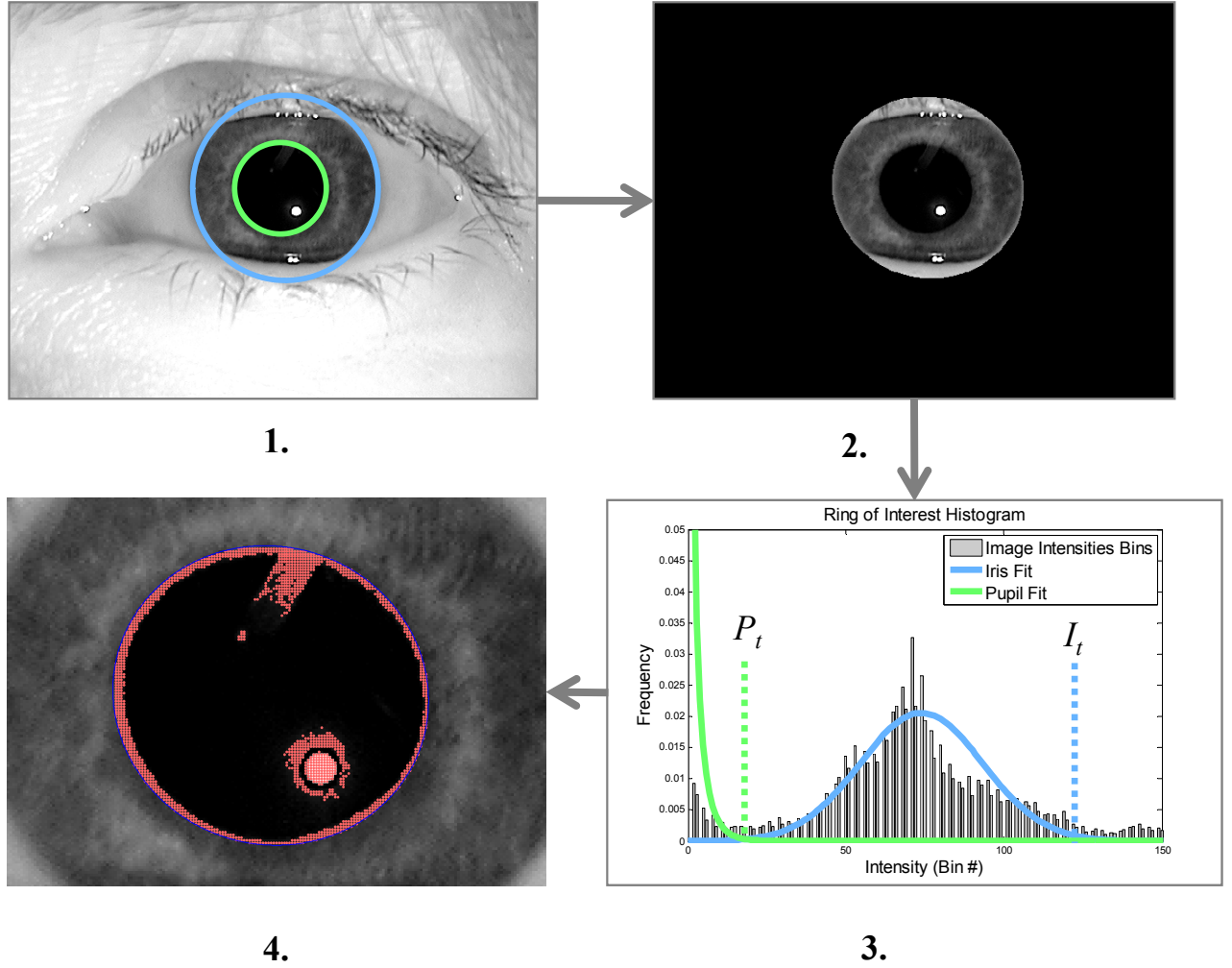


Figure 4.3: (1.) An iris image and its overlaid segmentation boundaries. (2.) Mask out all regions outside of the segmentation. (3.) Compute the intensity histogram of the unmasked region and estimate parameters for each model. For this specific image, $\text{Bin}(P_t) = 24$, $\text{Bin}(I_t) = 123$, $\hat{\Theta} = 4.75$, $\hat{\mu} = 73.93$, and $\hat{\sigma} = 19.39$. (4.) Every pixel within the estimated pupil segmentation is classified as belonging to the pupil or not. Red pixels were classified as non-pupil pixels; blue circle corresponds to the estimated pupil segmentation boundary from Masek's algorithm.

employ a Gaussian distribution for $P(\bar{x}|H_0) \sim N(\mu, \sigma^2)$.

The last point of concern is estimating parameters for each model, which are computed

“online” (we assume the shape parameter $k = 1$ for all experiments thus only the scale parameter Θ is estimated for the Gamma model). That is, for each image, a new set of parameters is estimated for each model. This is accomplished as follows. First, the image region of interest (ROI) is localized (i.e. removing everything but the segmentation result). The spatial histogram of the image intensities for the ROI is computed. Once the histogram has been computed, the scale parameter is estimated as:

$$\hat{\Theta} = \sum_{i=0}^{Bin(P_t)} x_i w_i, \quad (4.5)$$

where P_t is a threshold used to constrain the size of the pupil region, x_i is a gray level bin from the histogram of the ROI, and w_i is the weight associated with bin x_i which sums to one (the normalizing term $\frac{1}{k}$ is omitted since we assume $k = 1$). In other words, $\hat{\Theta}$ is derived by summing the product of the gray level bins and the associated weights until the bin corresponding to P_t is reached. Similarly, the parameters for the Gaussian are estimated as:

$$\hat{\mu} = \sum_{i=Bin(P_t)+1}^{Bin(I_t)} x_i w_i, \quad \hat{\sigma}^2 = \sum_{i=Bin(P_t)+1}^{Bin(I_t)} w_i (x_i - \hat{\mu})^2, \quad (4.6)$$

where I_t is a threshold used to constrain the size of the iris region, x_i is a gray level bin from the histogram of the ROI, and w_i is the weight associated with bin x_i which sums to one. Thresholds, P_t and I_t , have been determined based on experimental evaluation (all results presented with $P_t = 10000$ and $I_t = 25000$). Finally, every pixel within the pupil boundary is assigned 0 or 1 based on equation 4.4 and the ratio of these values is used as the pupil

over-segmentation score. Fig. 4.3 is a block diagram of this process.

The described metric is designed to measure pupil over-segmentation, P_{over} , that is when the estimated pupil boundary is greater than the actual pupil boundary. Pupil under-segmentation, when the estimated pupil boundary is smaller than the actual boundary, would remain undetected. To accommodate under-segmentation, we employ an iterative approach that increases the estimated pupil radius (or in the case of an ellipse, the estimated major and minor axis) and determines whether those pixels inside the expanded pupil radius are pupil pixels by using equation (4.4). This process continues until the pupil radius has reached the size of the iris radius or the ratio of pupil to non-pupil pixels is less than 20% which was chosen based on experimental evaluation. The rationale behind the addition of the threshold was to prevent the influence of heterogenous factors such as dark eyelashes/eyelids and reduce unnecessary computations when the estimated pupil boundary is not under-segmented. The final under-segmentation score is calculated as:

$$P_{under} = \frac{P_{over}}{P_{over} + P_{est_under}}. \quad (4.7)$$

The over-segmentation score is utilized here because the pupil boundary may contain non-pupil pixels whereas P_{over} is an estimate of just the pupil pixels within the pupil boundary. P_{est_under} is the total number of estimated pupil pixels over all iterations (displayed as green in Fig 4.4(b)). Figure 4.4 is an illustration of this process. Figure 4.4(a) is a failed pupil and iris segmentation. The blue ellipse is the estimate of the pupil boundary while the red ellipse is the estimated iris boundary. Our pupil over-segmentation produces a score of $P_{over} = 0.74$ indicating that there is no over-segmentation. Figure 4.4(b) is an illustration of the masked

region overlayed with the estimated pupil pixels (green) and increasing elliptical bands (red). Our pupil under-segmentation score for this image is $P_{under} = 0.44$.

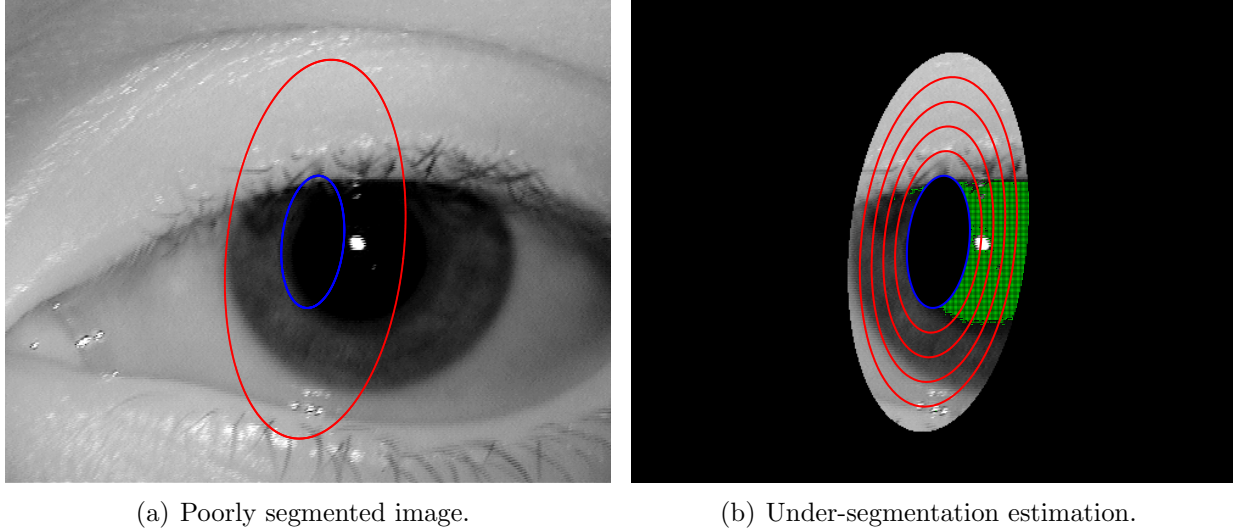


Figure 4.4: Pupil under-segmentation estimation (a) Under segmented pupil (b) Estimated under-segmentation.

4.3.3 Edge Measure

The last type of feature utilized for segmentation evaluation is related to edge information. Iris segmentation boundaries fall along a region edge such as the iris/pupil and iris/sclera borders. A simple, but effective measure should evaluate whether there is sufficient edge information along the estimated boundary. Furthermore, edge features should provide more information regarding the precision of the segmentation. Unfortunately, under non-ideal conditions, such as the introduction of blur or poor illumination, edges become distorted such that the intensity change is much more gradual, ultimately making detection more difficult. In order to help mitigate this problem, we propose a measure that is constrained

to specific sampling angles along the boundary but also in the radial direction of the search. More specifically, consider a radial slice, R , of image intensities of length N . This slice is centered on, for instance, the iris boundary at sampling angle θ , counter-clockwise from the iris center. We wish to compute edge information and evaluate whether it is near the center of the estimated boundary (e.g., $\frac{N}{2}$ within the radial slice). Edge information is estimated by summing the intensities within the radial slice over a window of size, w . Finite differences, \overline{D}_θ , are computed from the sums along the entire length of the radial slice as follows:

$$\overline{D}_\theta = \sum_{i=k-w}^{k-1} R_i - \sum_{i=k+1}^{k+w} R_i. \quad (4.8)$$

Next, finding the minimum difference in, \overline{D}_θ , will yield the location within the radial slice that corresponds to the region edge, according to equation 4.8. Note that we are also enforcing, by lack of absolute difference operator in equation 4.8, the assumption that the pupil region contains darker intensities than the iris region and that the iris region has darker intensities than the sclera. This effectively biases the operator towards region boundaries that transition from darker to lighter intensities. Finally, the Euclidean distance is computed between the estimated region edge location found by equation 4.8, and the iris boundary point at angle θ . The distance value is normalized by $\frac{N}{2}$ since the region edge can be inside the iris boundary, outside the iris boundary, or in the case of correct segmentation, overlaying the iris boundary. After normalization between $([0,1])$, the value is complemented, which serves as an informal score for characterizing edge information. Edge scores approaching 1 indicate correct boundary localization while scores approaching 0 indicate the opposite.

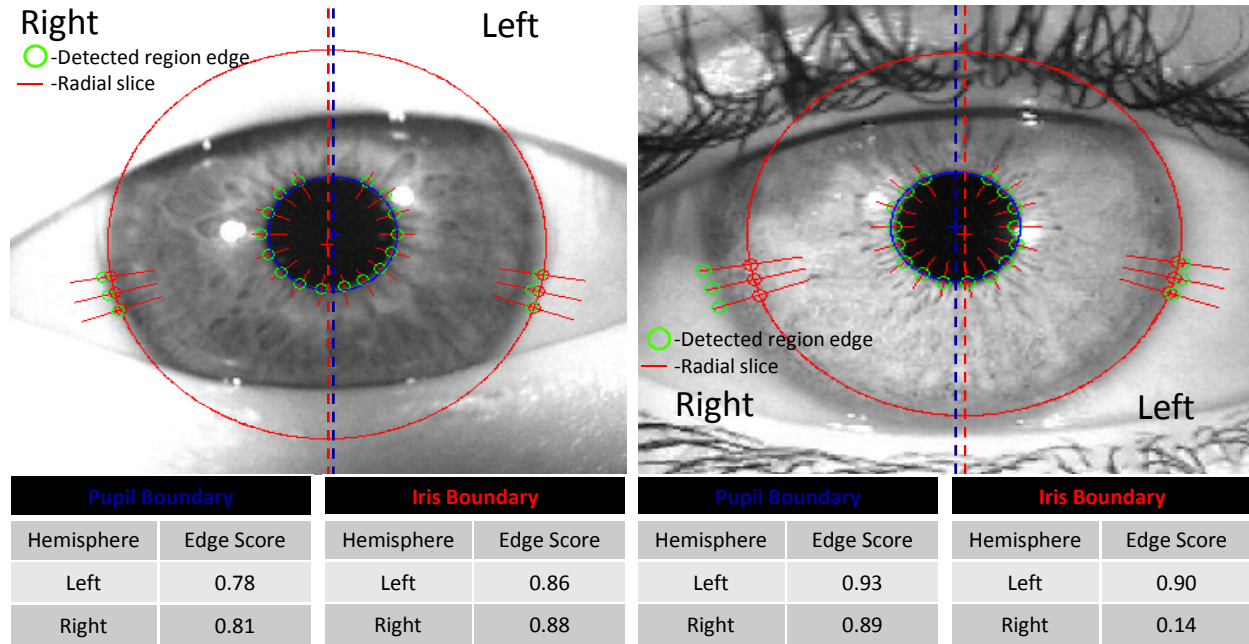


Figure 4.5: Illustration of the region edge measure. The left column represents a correct segmentation for both pupil and iris boundaries while the image in the right column has an inferior iris boundary. The red circles represent points on the estimated pupil/iris boundary at specific sampling angles. The iris boundary sampling angles are 10,15,20 for the left hemisphere and 170,165,160 for the right hemisphere in both images. The green circles correspond to the location of the edge region detected by equation 4.8. Note that the tables below the images represent the average edge scores for the sampling angles on the left and right hemisphere of the boundary respectively.

4.4 Experimental Design

To test the proposed technique, we selected three publicly available iris data sets and three iris recognition algorithms (including segmentation, encoding, and matching blocks). For data sets, we chose the WVU non-ideal iris set [90], the NIST Iris Challenge Evaluation (ICE) set [78], and an extremely challenging subset of imagery from Q-FIRE [93]. It is important to note that an overwhelming majority of imagery in this subset were attenuated by blur and off-angle. While commercial systems employ quality control algorithms similar

to the descriptions in Chapter 3, there is still value in keeping such imagery to evaluate the proposed segmentation evaluation methodology. In particular, this subset was extracted from the subject videos corresponding to low illumination/blur and high illumination/angles at 5ft for all 90 subjects. Four frames of the right iris were extracted from each video at near equal spaced intervals (e.g., frame indexes 23, 38, 76, and 113). Each frame was then canonicalized to a size of 480x640 such that the iris region was dominating the center of the image region (as in [100]). Each image was then manually inspected. Image frames where the iris was completely occluded by eyelids/eyelashes, out of the frame entirely (e.g., the iris was completely outside the original video frame as in 4.6(d)), or blurred to the point where the pupil or iris is unrecognizable to a human observer, were discarded. Thus, at the time

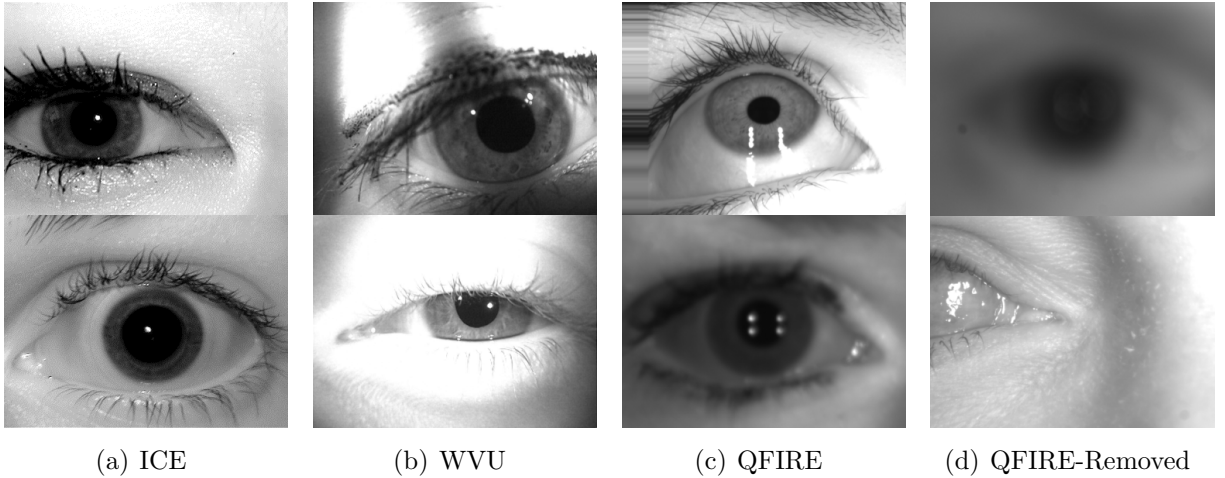


Figure 4.6: Sample imagery from each data set: (a) ICE, (b) WVU, and (c) QFIRE. Images discarded during frame extraction of the QFIRE videos are illustrated in (d). It is important to note that all three data sets are not indicative of operational data because quality control was not employed during collection. Thus, they are of particular interest for the focus of this work, on processing strategies for non-ideal iris.

of these experiments, the WVU data contained 2,412, the ICE data set contained 2,953

images, and Q-FIRE subset contained 583 images.

For each data set, all images were segmented with three segmentation algorithms. The first algorithm was a WVU in-house algorithm authored by Zuo et. al. [130]. The second is Masek’s publicly available implementation [65] while the third is our own implementation of Daugman’s integro-differential operator [27]. It is worthy of noting that we did not optimize the segmentation parameters of any of the algorithms to the individual data sets. After having segmented both data sets with both algorithms, the segmentation results were ground truthed. Figure 4.1 shows example segmentations and the caption lists associated ground truth results. Table 4.1 summarizes the segmentation ground truth results from all combinations of algorithms and data sets. Note that ground truthing was liberally enforced. If an estimated boundary did not completely delineate the intended (e.g., iris/pupil or iris/sclera) region, then it was deemed failed. Annotation was performed by visual evaluation of each boundary.

Category	Segmentation Failures								
	Zuo et al.			Masek			IDO		
	WVU	ICE	Q-FIRE	WVU	ICE	Q-FIRE	WVU	ICE	Q-FIRE
Pupil	23 (0.95%)	19 (0.64%)	93 (15.95%)	146 (6.02%)	185 (5.28%)	352 (60.38%)	110 (4.56%)	100 (3.39%)	94 (16.12%)
Iris	31 (1.29%)	23 (0.70%)	90 (15.44%)	858 (33.20%)	214 (6.60%)	293 (50.26%)	369 (15.22%)	85 (2.88%)	102 (17.50%)
Either	37 (1.53%)	24 (0.81%)	106 (18.18%)	923 (35.96%)	379 (11.24%)	423 (72.56%)	408 (16.92%)	135 (4.57%)	136 (23.33%)

Table 4.1: Segmentation ground truth results for all algorithms across WVU, ICE, and QFIRE data sets. Note that the Zuo et al. algorithm failed to produce segmentation boundaries for 13 of the 583 in the QFIRE subset, which were excluded from subsequent analysis.

Given the subjective nature of ground truthing, we provide histograms of Tanimoto coefficients, T , between the ground truth data and the automatic segmentation algorithms for the ICE data set only as this requires manual annotation of both boundaries. Mathematically,

calculation of the Tanimoto coefficient is described as:

$$T(S_g, S_a) = \frac{N_{ga}}{N_g + N_a - N_{ga}}, \quad (4.9)$$

where S_g and S_a correspond to segmentation produced from the ground truth and automatic segmentation algorithms. The number of unmasked pixels produced by the intersection of both segmentation masks is represented by N_{ga} , while N_g and N_a , correspond to unmasked pixels from their respective segmentations. The average Tanimoto coefficient(s) for correct segmentation are 0.95 for both Masek and IDO algorithms respectively, while for failed seg-

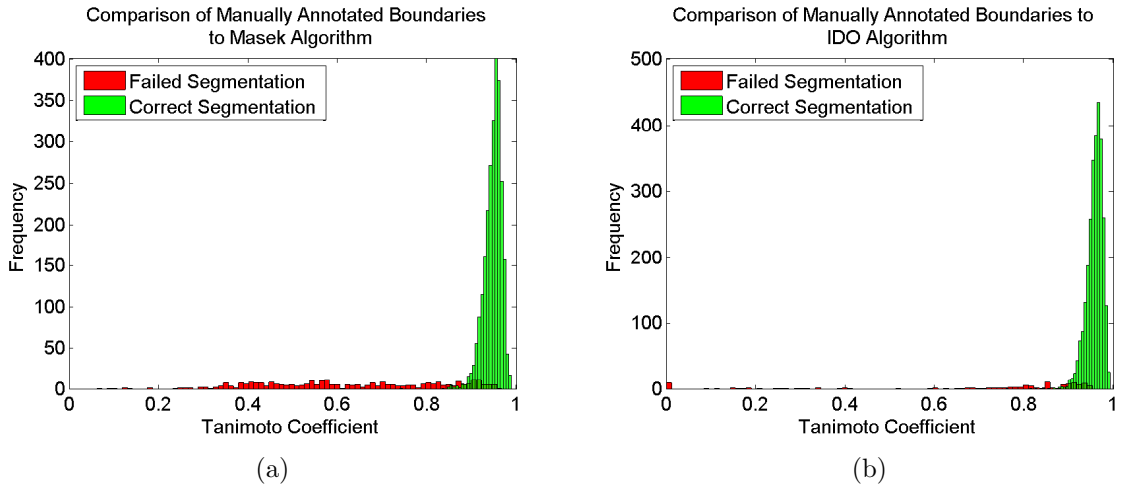


Figure 4.7: Histograms of Tanimoto coefficient for both Masek and IDO algorithms. The Tanimoto coefficient, in this context, is utilized to characterize the ground truth for each segmentation algorithm. That is, ground truth for correct segmentation on average correspond to a coefficient of $.95$. On the other hand, segmentation below a coefficient of $.95$ correspond to failed segmentation.

mentation the averages are 0.62 and 0.68 respectively. The Zuo et al. algorithm has a mean Tanimoto coefficient of 0.94 and $.46$ for correct and failed segmentation respectively. Plot 4.7 illustrates the Tanimoto histograms for both Masek and IDO segmentation algorithms.

These plots are useful in that they can be used to characterize how manual ground truth annotation compares to the automatic for both failed and correct segmentation. Throughout the rest of this dissertation, we may refer to failed segmentation as inferior, or unfavorable. Likewise, we may refer to correct segmentation boundaries as good, favorable or accurate.

After the ground truth results were tabulated, we processed the segmented images for the nine combinations of algorithms and data sets by scoring the pupil and iris segmentation results with the technique described in the following section. After processing the images, we ran the results, including pupil and iris scoring, through a ensemble of machine learners to arrive at a simple binary segmentation result. That is, the model evaluates that either both the pupil and iris boundaries were correctly estimated (good segmentation) or at least one of the boundaries was incorrectly estimated (failed segmentation). Finally, we look into the effect that varying iris segmentation results has on iris match scores by filtering the match score results according to the model evaluations.

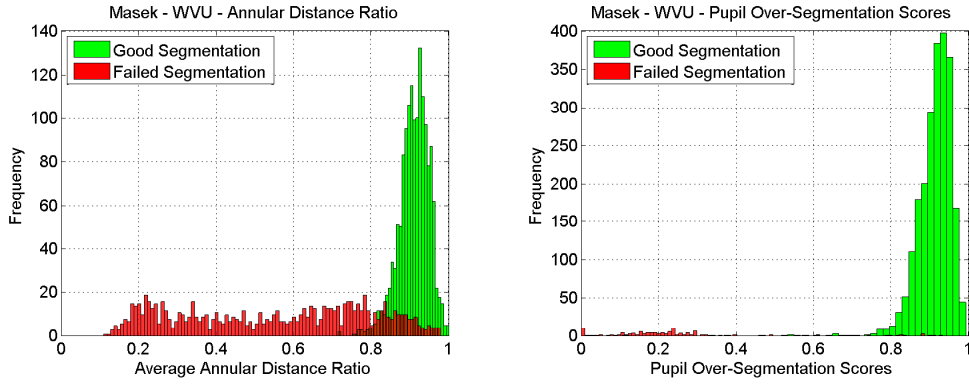
4.5 Experimental Results

This section is broken down into five subsections. First, we establish the ability of the described scoring approach to accurately evaluate segmentation results for both pupil and iris segmentation. Next, we present the ability of the ensemble to evaluate overall segmentation results. Then to demonstrate one potential application of the technique, we present the effect of filtering iris recognition match scores by the evaluated segmentation results. The primary goal of this chapter is to arrive at an automatic technique for measuring segmentation results of both pupil and iris segmentation boundaries. With that in mind, we tested the scoring

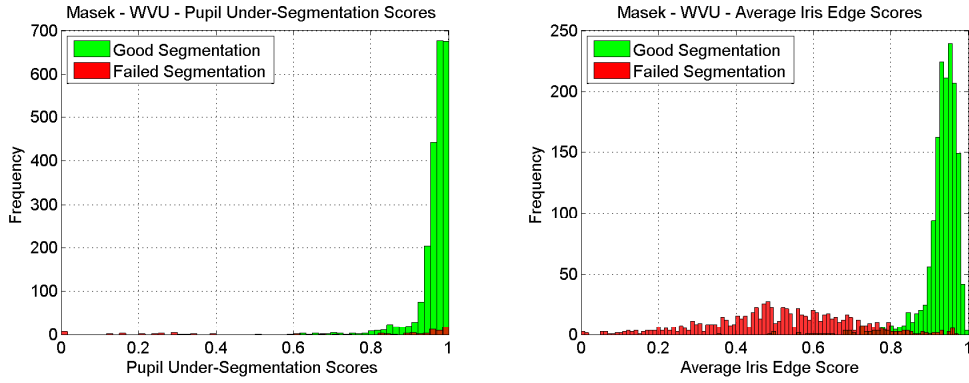
techniques on three data sets for three different segmentation algorithms.

4.5.1 Masek Segmentation

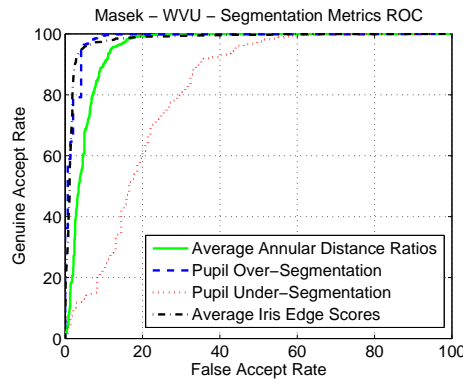
First we look at the pupil and iris scoring with the Masek algorithm and the WVU data set. Figure 4.8 shows the distributions of scores for images that have both good segmentation and failed segmentation. Figure 4.8(a) depicts the average annular distance ratio scoring results with the red histogram representing the distribution for images that failed segmentation and the green histogram corresponding to images that were correctly segmented. As can be seen by the plot, the distributions are separated fairly well with the correctly segmented images having a mean score of 0.91 and the images that failed pupil segmentation having a mean score of 0.55 . Figure 4.8(b) illustrates the pupil over-segmentation scores which have better separation between the distributions than the average annular histograms. The distributions for both failed and correct segmentation have a mean of 0.27 and 0.91 , respectively. On the other hand, the pupil under-segmentation scores shown in plot 4.8(c) have more overlap than the over-segmentation distributions. Here the scores corresponding to images with correctly segmented regions have a mean of 0.96 while the failed images scored 0.64 on average. Figure 4.8(d) provides the distributions for the average edge scores. Like the annular distance ratio distributions, there is fairly good separation between correct and failed segmentation. That is, the mean for correct segmentation is 0.93 while the mean for failed segmentation is 0.52 . Plot 4.8(e) displays GAR/FAR ROC curves for both the pupil and iris segmentation scoring. Here we define a “genuine accept” to be a correctly segmented image classified as correctly segmented and a “false accept” to be an incorrectly segmented image classified as correctly segmented. Based on the distributions seen in plots (a), (b), (c), and (d) it is not surprising



(a) Masek-WVU: Average Annular Distance Ratios. (b) Masek-WVU: Pupil Over-segmentation.



(c) Masek-WVU: Pupil Under-segmentation. (d) Masek-WVU: Average Iris Edge Scores.

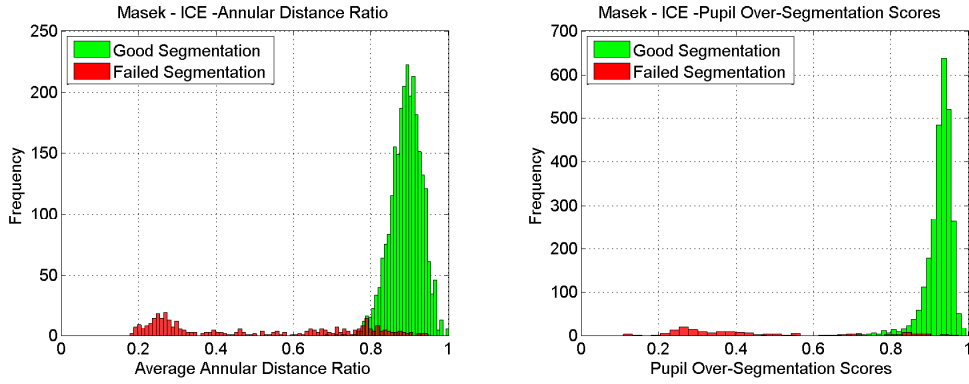


(e) Masek-WVU: Metric ROC.

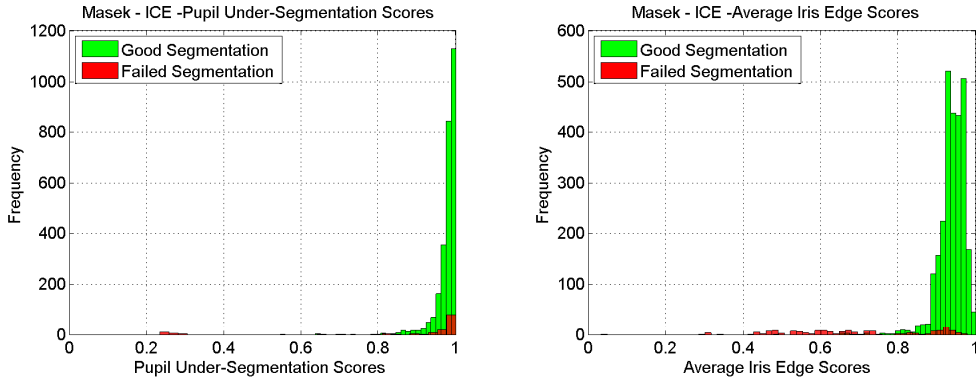
Figure 4.8: Segmentation score distributions for Masek segmentation on WVU data (a) Average annular distance ratio distributions (b) Pupil-Over distributions (c) Pupil-under distributions (d) Average iris edge score distributions (e) Segmentation metric ROC.

to observe that the average annular distance ratios, pupil over-segmentation, and average edge score ROCs indicate fairly good performance while the pupil under-segmentation curves illustrate the existence of errors with an $EER \approx 25.56\%$.

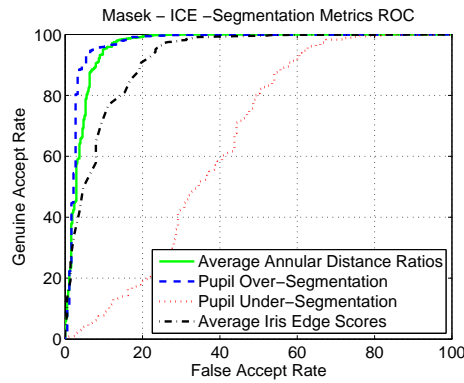
In Figure 4.9 we refer to evaluation of the Masek algorithm when processing the ICE data. Like Figure 4.8, distributions of scores for images that have both good segmentation and failed segmentation are represented by the green and red histograms respectively. Plot 4.9(a) illustrates the distributions for the average annular distance ratio scores that appear very similar to the ratio scores of the WVU data. Here the mean scores for correct and failed segmentation are 0.89 and 0.49 respectively. Likewise, the distributions for pupil over-segmentation, 4.9(b), are similar to the scores generated from the WVU data, with mean scores of 0.92 and 0.44 . On the other hand we see a preponderance of overlapping scores in the pupil under-segmentation, presented in plot 4.9(c). Here the mean pupil under-segmentation score for images that failed segmentation increased, 0.81 , as with the WVU scores, while the mean for correct segmentations, 0.97 , slightly increased. Thus, the overall separation of distributions for pupil under-segmentation decreased in comparison to the WVU scores. Following this trend, the distributions for the average iris edge scores, plotted in 4.9(d), also indicate a slight decrease in separation. Notably, the mean for correct and failed segmentation is 0.94 down from 0.67 respectively. As a result, we see the performance of the iris segmentation classification decrease in plot 4.9(e). More specifically, the EER increases to 15.90% , whereas with the WVU data, an EER of 4.56% was obtained. This is further observed as the ROC performance of the iris edge scores performed the best with the WVU data. Whereas, the pupil-over segmentation provides the best performance when evaluating the ICE data. Also worth mentioning, the ROC performance of the pupil under-



(a) Masek-ICE: Average Annular Distance Ratios. (b) Masek-ICE: Pupil Over-segmentation.



(c) Masek-ICE: Pupil Under-segmentation. (d) Masek-ICE: Average Iris Edge Scores.



(e) Masek-ICE: Metric ROC.

Figure 4.9: Segmentation score distributions for Masek segmentation on ICE data (a) Average annular distance ratio distributions (b) Pupil-Over distributions (c) Pupil-under distributions (d) Average iris edge score distributions (e) Segmentation metric ROC.

segmentation decreases, only attaining an EER of 40.57%.

Table 4.2 provides summary statistics of the evaluation measures when processing the Q-FIRE subset. Here we observe that the mean scores for correct segmentation are fairly well separated from those pertaining to failed segmentation. In particular, the mean scores for correct segmentation are .8868, .9553, .9276, and .8968 for annular ratio, pupil over-segmentation, pupil under-segmentation, and iris edge information respectively. In contrast, the mean scores for failed segmentation are .3367, .2559, .7389, and .5649 for annular ratio, pupil over-segmentation, pupil under-segmentation, and iris edge information respectively. Clearly, these features provide information that can be useful for discriminating between correct and incorrect segmentation boundaries. Note that there is more overlap for the pupil under-segmentation metric scores, indicating that the characterization of failed segmentation boundaries for this algorithm and data set combination is influenced more from pupil over-segmentation and iris boundary failures. This is also observed in Fig. 4.10, which illustrates ROC performance for all feature types on the QFIRE data set. The pupil under-segmentation feature provides inferior performance in comparison to the other features.

4.5.2 IDO Segmentation

Next, we look at the pupil and iris scoring with our implementation of the integro-differential operator algorithm and the WVU data set. Figure 4.11 shows the distributions of scores for images that have both good segmentation and failed segmentation. Figure 4.11(a) depicts the average annular distance ratio scoring results with the red histogram representing the distribution for images that failed segmentation and the green histogram corresponding to images that were correctly segmented. As illustrated by the plot, the distributions completely

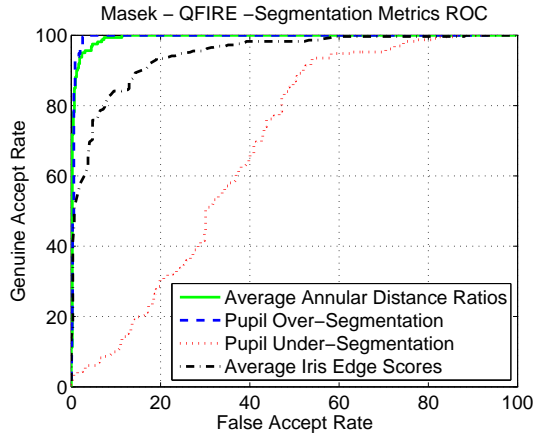


Figure 4.10: ROC performance for the segmentation metrics generated from the Masek segmentation boundaries on the QFIRE data set.

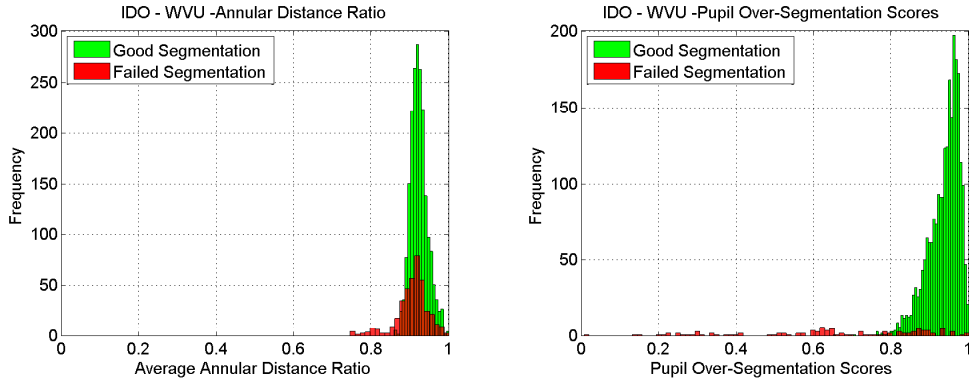
Segmentation Metrics	Mean	Std
Failed Annular Ratio	.3367	.2901
Correct Annular Ratio	.8868	.0862
Failed Pupil-Over	.2559	.1668
Correct Pupil-Over	.9553	.0550
Failed Pupil-Under	.7389	.3028
Correct Pupil-Under	.9276	.0882
Failed Iris Edge	.5649	.3672
Correct Iris Edge	.8968	.1599

Table 4.2: Segmentation metric distribution statistics for Masek segmentation on the QFIRE data set.

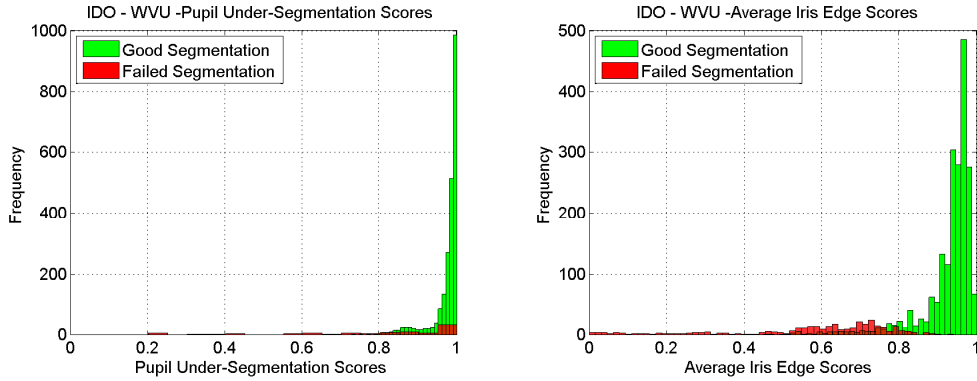
overlap with the correctly segmented images having a mean score of 0.92 and the images that failed segmentation having a mean score of 0.90 . In contrast, the same features for the Masek algorithm were well separated. Figure 4.11(b) presents the pupil over-segmentation scores which have much better separation between the distributions than the average annular histograms. The distributions for both failed and correct segmentation have a mean of 0.64 and 0.94 , respectively. On the other hand, the pupil under-segmentation scores shown in plot 4.11(c) have more overlap than the over-segmentation distributions. Here the scores corresponding to images with correctly segmented regions have a mean of 0.97 while the failed images scored 0.75 on average. The iris edge score distributions are provided in plot 4.11(d). Notably, the distributions are fairly well separated with a mean of 0.93 for correct iris segmentation and 0.58 for failed iris segmentation. This is corroborated by plot 4.11(e) which illustrates the ROC performance of all metric scores for the IDO algorithm on the WVU data set. More specifically, the EER for the iris edge scores, the black line

plot, is 6.87% . In contrast, the EERs for the remaining metrics are inferior to the iris edge metric. That is, the EERs for annular distance ratio, pupil over-segmentation, and pupil under-segmentation are 41.62% , 14.55% , and 25.50% respectively.

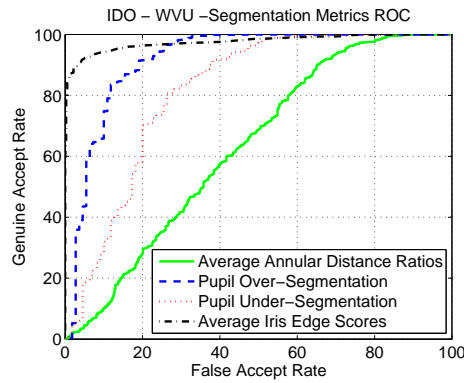
Figure 4.12 shows the distributions of scores for images that have both good segmentation and failed segmentation for the ICE data set when processed with the IDO segmentation algorithm. Figure 4.12(a) depicts the average annular distance ratio scoring results with the red histogram representing the distribution for images that failed segmentation and the green histogram corresponding to images that were correctly segmented. As illustrated by the plot, the distributions completely overlap with the correctly segmented images having a mean score of 0.91 and the images that failed segmentation having a mean score of 0.88 . This overlap was also observed for scores generated from the WVU data set. Figure 4.12(b) presents the pupil over-segmentation scores which have much better separation between the distributions than the average annular histograms. The distributions for both failed and correct segmentation have a mean of 0.67 and 0.93 , respectively. On the other hand, the pupil under-segmentation scores shown in plot 4.12(c) have more overlap than the over-segmentation distributions. Here the scores corresponding to images with correctly segmented regions have a mean of 0.97 while the failed images scored 0.83 on average. The iris edge score distributions are provided in plot 4.12(d). The distributions are fairly well separated with a mean of 0.95 for correct iris segmentation and 0.61 for failed iris segmentation. This trend was also observed for scores generated from the WVU data set. This is further corroborated by plot 4.12(e) which illustrates the ROC performance of all metric scores for the IDO algorithm on the ICE data set. More specifically, the EER for the iris edge scores, black line plot, is 3.51% . Likewise, the EER for the pupil over-segmentation scores is



(a) IDO-WVU: Average Annular Distance Ratios. (b) IDO-WVU: Pupil Over-segmentation.



(c) IDO-WVU: Pupil Under-segmentation. (d) IDO-WVU: Average Iris Edge Scores.



(e) IDO-WVU: Metric ROC.

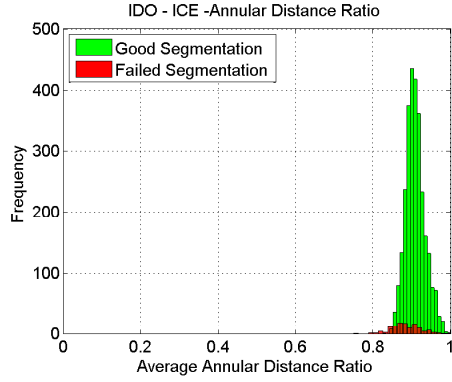
Figure 4.11: Segmentation score distributions for IDO segmentation on ICE data (a) Average annular distance ratio distributions (b) Pupil-Over distributions (c) Pupil-under distributions (d) Average iris edge score distributions (e) Segmentation metric ROC.

also fairly reasonable at 4.00% . Similar to performance of the WVU data set, the remaining EERs of the other measures are inferior to the pupil over-segmentation and iris edge measures. Specifically, the EERs for annular distance ratio and pupil under-segmentation are 36.28% and 31.05% respectively.

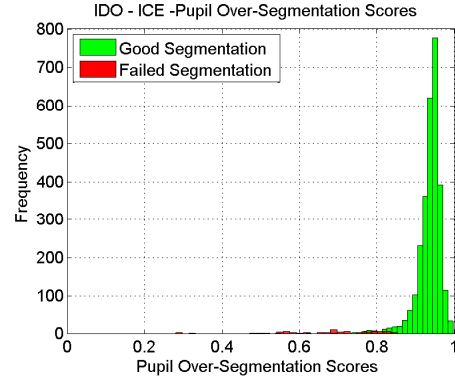
In Table 4.3 we provide summary statistics for both correct and failed IDO segmentation when processing the subset of imagery from the Q-FIRE data set. The more notable observation in comparison to the statistics provided for the Masek algorithm, is that we less separation between the correct and failed distributions for the annular distance ratio scores. The mean of correct segmentation for annular distance ratios are $.9076$ while for failed segmentation the means shift to $.8907$. This observation holds across all data sets for this measure when evaluating segmentation from the IDO algorithm. The cause likely results from our implementation of the algorithm. After finding the pupil boundary, the iris search is constrained to evaluate locations within the pupil region near the pupil center, thus limiting the potential for gross inaccuracies. On the other hand, we observe slightly more separation for the iris edge information measure. Specifically, a mean of $.8879$ for correct segmentation is attained while for failed segmentation the mean is $.5222$.

4.5.3 Zuo et al. Segmentation

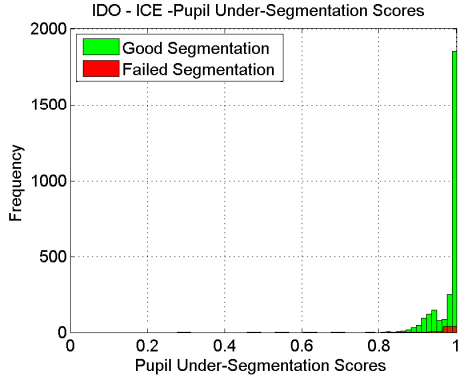
As noted in the experimental design, Zuo’s algorithm makes far fewer errors across all data sets in comparison to the other segmentation algorithms. As a result, the distributions for the failed annular ratio, pupil-over, pupil-under, and iris edge scores are not suitable for graphical representation. However, Table 4.4 characterizes the data in terms of mean and standard deviation. Here the mean annular ratio scores fall at 0.9310 and 0.8636 for correct and failed



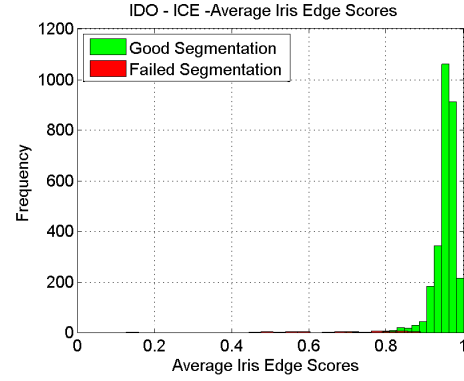
(a) IDO-ICE: Average Annular Distance Ratios.



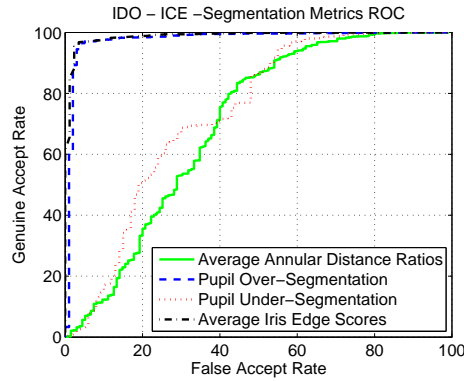
(b) IDO-ICE: Pupil Over-segmentation.



(c) IDO-ICE: Pupil Under-segmentation.



(d) IDO-ICE: Average Iris Edge Scores.



(e) IDO-ICE: Metric ROC.

Figure 4.12: Segmentation score distributions for IDO segmentation on ICE data (a) Average annular distance ratio distributions (b) Pupil-Over distributions (c) Pupil-under distributions (d) Average iris edge score distributions (e) Segmentation metric ROC.

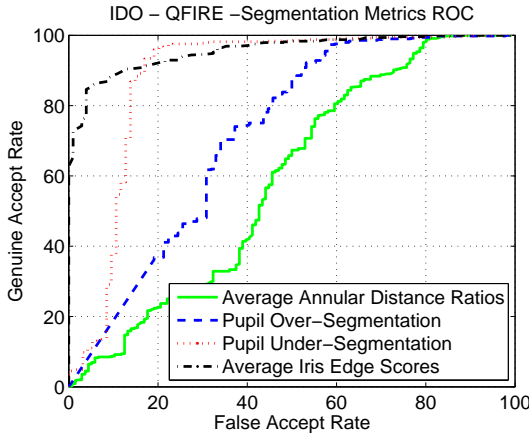


Figure 4.13: ROC performance for the segmentation metrics generated from the IDO segmentation boundaries on the QFIRE data set.

Segmentation Metrics	Mean	Std
Failed Annular Ratio	.8907	.0716
Correct Annular Ratio	.9076	.0518
Failed Pupil-Over	.7467	.2702
Correct Pupil-Over	.9486	.0892
Failed Pupil-Under	.5702	.2353
Correct Pupil-Under	.9285	.0937
Failed Iris Edge	.5222	.3663
Correct Iris Edge	.8879	.1852

Table 4.3: Segmentation metric distribution statistics for IDO segmentation on QFIRE.

WVU segmentation respectively. This is consistent with the scores for the IDO segmentation algorithm but inconsistent with the Masek algorithm. Similarly, the pupil-over and iris edge segmentation score distributions are consistent with the IDO algorithm, corresponding to a mean of 0.9500 and 0.9018 for correctly segmented images respectively. This

Segmentation Metrics	ICE		WVU		Q-FIRE	
	Mean	Std	Mean	Std	Mean	Std
Failed Annular Ratio	0.8004	0.1166	0.8636	0.0775	.6025	.20352
Correct Annular Ratio	0.9123	0.0351	0.9310	0.0307	.8829	.0839
Failed Pupil-Over	0.6099	0.2696	0.6052	0.1792	.5637	.3063
Correct Pupil-Over	0.9638	0.0386	0.9500	0.0502	.9960	.0243
Failed Pupil-Under	0.7438	0.2972	0.8413	0.2552	.5541	.2730
Correct Pupil-Under	0.9789	0.0407	0.9626	0.0499	.9390	.0641
Failed Iris Edge	0.6403	0.2895	0.6854	0.1508	.5179	.3331
Correct Iris Edge	0.9251	0.0417	0.9018	0.0805	.8770	.1937

Table 4.4: Segmentation metric distribution statistics for Zuo's segmentation.

observation is not unexpected given the fact that part of the methodology behind the Zuo

et al. segmentation algorithm is comprised of the IDO operator (e.g., elliptical IDO) [130]. Looking at the segmentation metric scoring on the ICE data, Figure 4.14(c), with Zuo et al.'s segmentation algorithm, we see that the pupil-over segmentation distributions have a mean of 0.9638 and $.6099$ for correct and failed segmentations respectively. The ROC performance for pupil-over and iris edge metric scores generated from the WVU data, illustrated in Figure 4.14(b), as a prediction mechanism is similar to what we see in the Masek-WVU, IDO-WVU, and IDO-ICE ROC results. More specifically, the pupil-over segmentations provide the best ROC performance, attaining an EER of 0.71% while the EER for annular distance ratio, pupil-under, and iris edge scores are 32.41% , 46.56% , and 15.00% respectively.

Figure 4.14(c) illustrates the ROC metric performance for the Zuo et al. algorithm on the

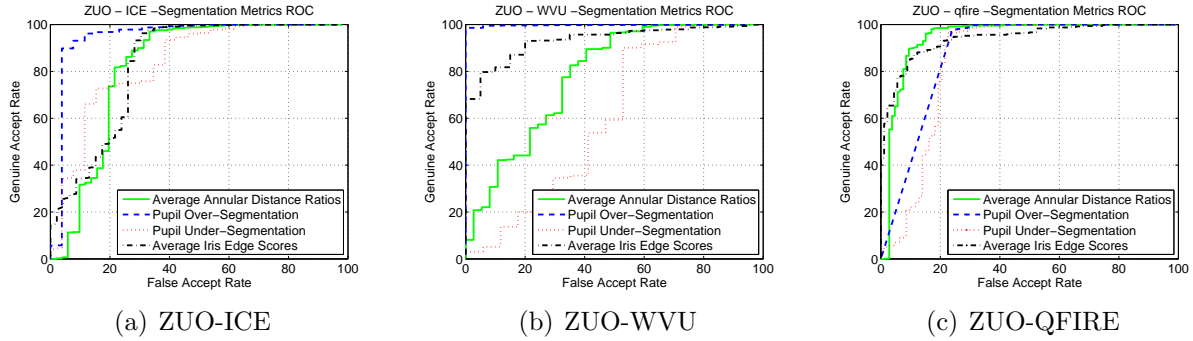


Figure 4.14: Pupil / iris ROC curves for Zuo et. al segmentation (a) ICE, (b) WVU, and (c) QFIRE.

QFIRE data. Interestingly, the annular distance ratio and iris edge metrics provide the best performance with EERs of 10.25% and 12.26% respectively.

4.5.4 Evaluating the Overall Segmentation Result

While the previous results are useful for measuring and subsequently evaluating the outcome of pupil and iris segmentation results independently, we have yet to explore the notion of evaluating overall segmentation success. That is, given the score(s) for annular distance ratio, pupil over/under - segmentation and iris edge metrics, can we evaluate whether both boundaries have been successfully segmented as in Figure 4.1(a)? Conversely, can we evaluate failed overall segmentation when the pupil, iris, or both boundaries are incorrectly estimated? As mentioned in the experimental design section, we chose to use a sequential ensemble of C-SVM [17] models, generated from each feature type. A block flow diagram

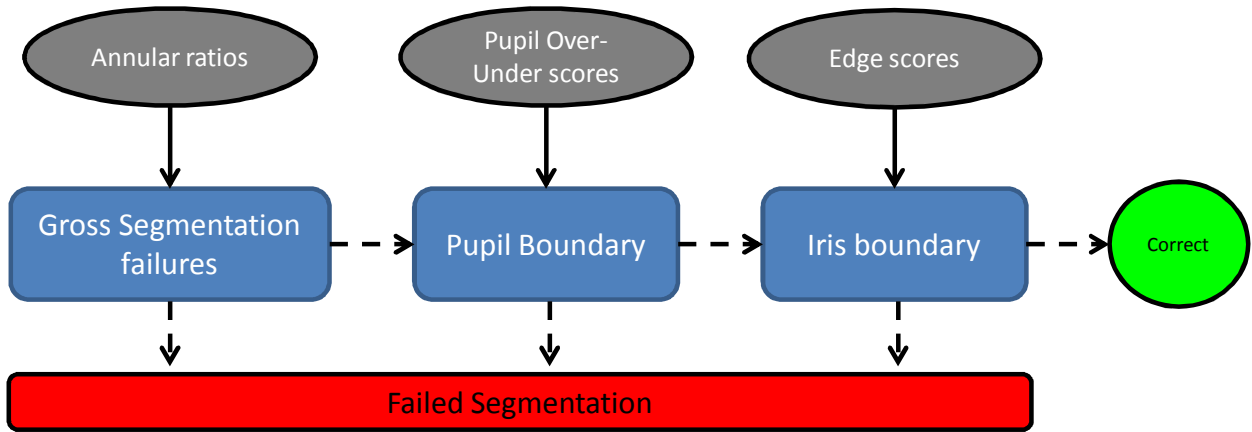


Figure 4.15: Illustration of the sequential C-SVM ensemble. Gross segmentation failures are detected with the model generated from annular distance ratios. If evaluated to be grossly inaccurate processing stops, otherwise the pupil features are evaluated within the next model. This process continues sequentially until the iris edge features are evaluated. If all models in the ensemble indicate successful segmentation, then the overall segmentation is classified as correct. Note that the dotted lines indicate optional processing steps since any model in the ensemble could detect an incorrect boundary, thus processing stops and subsequent models within the ensemble are not evaluated.

of the ensemble is illustrated in Figure 4.15. Due to the observation that other learning

approaches performed similarly well and SVMs have been extensively used in the biometric literature, we omit discussion on this specific machine learning technique. What is of interest to us is the performance of the two-class decision problem. Specifically, given the annular ratio, pupil over/under and iris edge score(s), how well does the C-SVM ensemble (or any other model of interest) evaluate the overall segmentation result?

Table 4.5 shows the confusion matrices for all combinations of segmentation algorithms and data sets utilizing just the annular distance ratio measure. It should be noted that only one parameter was changed from default for the WEKA implementation of the C-SVM algorithm. Notably, the normalize option was set to true. Additionally, 10x10 cross validation was used to train / test the models. We observe that with an ensemble containing just the annular distance ratio model, classification performance is only respectable for Masek and Zuo et al. (88.74% QFIRE only) algorithms, attaining average classification accuracies (average of the diagonal) of 91.37%, 92.51% and 94.29% for WVU, ICE, and QFIRE respectively. In contrast, the average classification accuracies for the IDO and Zuo et al. algorithms are inferior, achieving only 52.82%, 51.48%, 53.54%, 64.78%, and 73.98% for IDO-WVU, IDO-ICE, IDO-QFIRE, Zuo-WVU, and Zuo-ICE, respectively.

Table 4.6 illustrates classification performance when we add another C-SVM model, generated and trained with the pupil over/under-segmentation measures, to the ensemble. Here, we observe only a slight increase in average classification performance for the Masek algorithm but a modest increase for the others. More specifically, the average classification accuracies for Masek-WVU, Masek-ICE, Masek-QFIRE, IDO-WVU, IDO-ICE, IDO-QFIRE, Zuo-WVU, Zuo-ICE, and Zuo-QFIRE are 91.39%, 93.57%, 92.46%, 62.66%, 81.50%, 84.30%, 89.99%, 87.61% and 92.05% respectively. However, we also observe an increase in

(a)			(b)			(c)		
Masek - WVU			IDO - WVU			Zuo et. al - WVU		
	P G	P F		P G	P F		P G	P F
A G	98.33%	1.67%	A G	100%	0%	A G	99.83%	0.17%
A F	15.60%	84.40%	A F	94.36%	5.64%	A F	70.27%	29.73%

(d)			(e)			(f)		
Masek - ICE			IDO - ICE			Zuo et. al - ICE		
	P G	P F		P G	P F		P G	P F
A G	97.67%	2.33%	A G	100%	0%	A G	98.93%	1.07%
A F	12.66%	87.34%	A F	97.04%	2.96%	A F	50.98%	49.02%

(g)			(h)			(i)		
Masek - QFIRE			IDO - QFIRE			Zuo et. al - QFIRE		
	P G	P F		P G	P F		P G	P F
A G	90.00%	10.00%	A G	100%	0%	A G	88.79%	11.21%
A F	1.42%	98.58%	A F	92.65%	7.35%	A F	11.32%	88.68%

Table 4.5: Annular Distance Ratios: C-SVM Ensemble Misclassifications. A G = actual good segmentation, A F = actual failed segmentation, P G = predicted good segmentation, P F = predicted failed segmentation

the error associated with classifying correct segmentation. In particular, the average error across all combinations of data sets and algorithms increases from 2.94% to 7.33% when the pupil feature model is added to the ensemble.

Finally, Table 4.7 provides classification performance when the third model, generated from iris edge scores, is added to the ensemble. As with the previous table we see a moderate increase in average classification performance for all combination of algorithms and data sets. It is also important to note that the increase in average classification performance is attributed mainly to better classification of failed segmentation when additional models are added to the ensemble. Similar to the addition of the pupil feature model, it is worth noting that the classification errors for classifying correct segmentation boundaries are observed to increase, varying with the combination of data set and segmentation algorithm, when the

(a)			(b)			(c)		
Masek - WVU			IDO - WVU			Zuo et. al - WVU		
	P G	P F		P G	P F		P G	P F
A G	96.00%	4.00%	A G	97.12%	2.88%	A G	96.21%	3.79%
A F	13.22%	86.78%	A F	71.81%	28.19%	A F	16.22%	83.78%

(d)			(e)			(f)		
Masek - ICE			IDO - ICE			Zuo et. al - ICE		
	P G	P F		P G	P F		P G	P F
A G	94.52%	5.48%	A G	94.12%	5.88%	A G	96.79%	3.21%
A F	7.39%	92.61%	A F	31.12%	68.88%	A F	21.51%	78.43%

(g)			(h)			(i)		
Masek - QFIRE			IDO - QFIRE			Zuo et. al - QFIRE		
	P G	P F		P G	P F		P G	P F
A G	85.63%	14.38%	A G	87.70%	12.30%	A G	85.99%	14.01%
A F	0.71%	99.29%	A F	19.12%	80.88%	A F	1.89%	98.11%

Table 4.6: Annular Distance Ratios + Pupil Over/Under-Segmentation: C-SVM Ensemble Misclassifications. A G = actual good segmentation, A F = actual failed segmentation, P G = predicted good segmentation, P F = predicted failed segmentation

iris edge model is added to the ensemble. More specifically, the average error associated with classification of correct segmentation boundaries increases from 7.33% to 9.11% across all combination of algorithms and data sets.

4.5.5 Filtering Iris Match Scores Utilizing Segmentation Evaluation

The last result we present demonstrates one potential application of the technique. We use the overall segmentation result, output from the ensemble, to selectively filter match scores from the three different data sets. Due to the small number of incorrectly segmented images in the WVU and ICE data sets, we do not present results when using the Zuo's segmentation algorithm. Instead we focus on the Masek and IDO segmentation. As mentioned in the

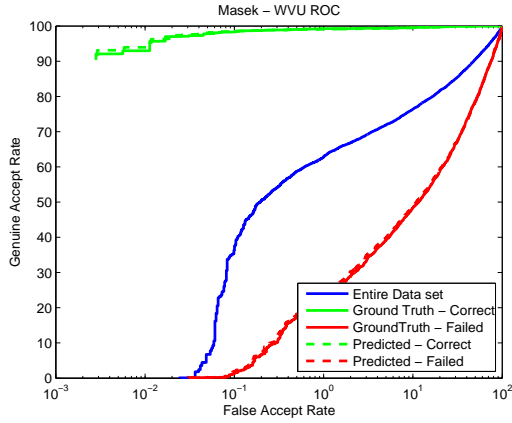
(a)			(b)			(c)		
Masek - WVU			IDO - WVU			Zuo et. al - WVU		
	P G	P F		P G	P F		P G	P F
A G	95.53%	4.47%	A G	94.64%	5.36%	A G	94.23%	5.77%
A F	1.52%	98.48%	A F	11.76%	88.34%	A F	8.11%	91.89%

(d)			(e)			(f)		
Masek - ICE			IDO - ICE			Zuo et. al - ICE		
	P G	P F		P G	P F		P G	P F
A G	94.17%	5.83%	A G	93.93%	6.07%	A G	96.52%	3.48%
A F	2.37%	97.63%	A F	7.40%	92.60%	A F	15.69%	84.31%

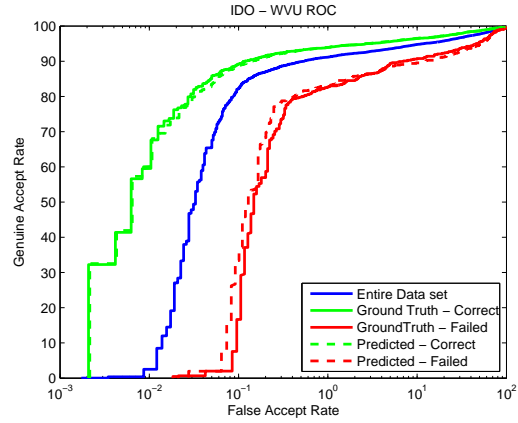
(g)			(h)			(i)		
Masek - QFIRE			IDO - QFIRE			Zuo et. al - QFIRE		
	P G	P F		P G	P F		P G	P F
A G	85.99%	15.63%	A G	82.33%	17.67%	A G	82.33%	17.67%
A F	0.24%	99.76%	A F	5.88%	94.12%	A F	1.89%	98.11%

Table 4.7: Annular Distance Ratio + Pupil Over/Under-Segmentation + Iris Edge: C-SVM Ensemble Misclassifications. A G = actual good segmentation, A F = actual failed segmentation, P G = predicted good segmentation, P F = predicted failed segmentation

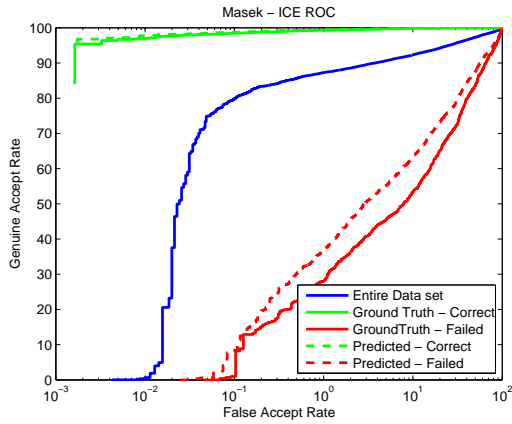
introduction, iris segmentation is a main factor in determining an iris recognition system's ability to successfully classify pairs of iris image as genuine or imposter. Along those lines, we would expect to see performance drop as the number of incorrectly segmented iris images increases. Figure 4.16 shows match score ROC curves based on segmentation results across the WVU, ICE and QFIRE data sets for both segmentation algorithms. In each Figure, a total of five curves are shown. The blue curves show the matching performance when all match scores are included and serves as a baseline. The highest performing solid (green) curves represent the matching performance from scores corresponding to pairs where both images were segmented correctly (ground truth). The highest performing dotted (green) curves represent the match scores of image pairs when the iris images were evaluated by the ensemble to have successfully segmented the iris image. Conversely, the lowest performing



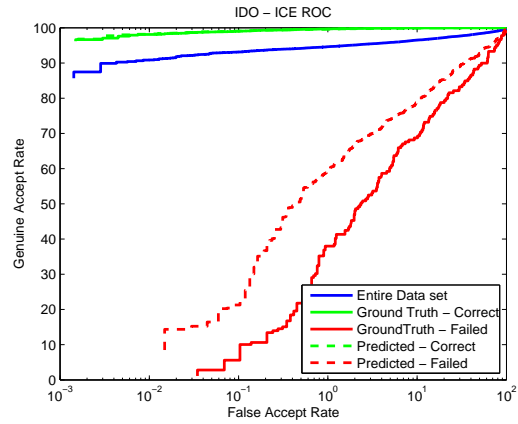
(a) Masek-WVU Iris Matching ROC.



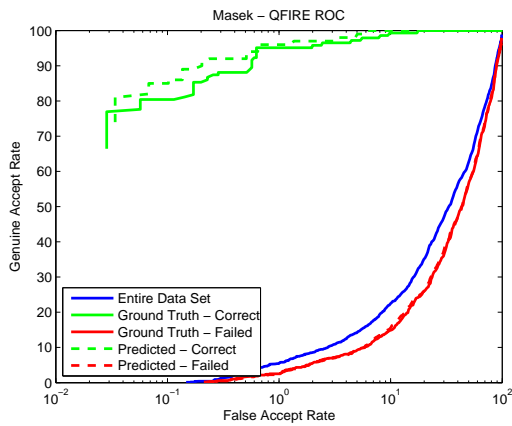
(b) IDO-WVU Iris Matching ROC.



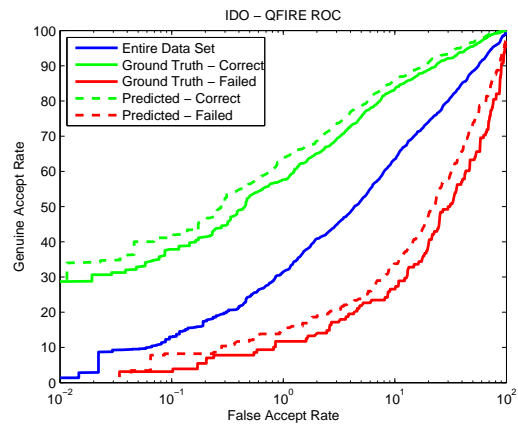
(c) Masek-ICE Iris Matching ROC.



(d) IDO-ICE Iris Matching ROC.



(e) Masek-QFIRE Iris Matching ROC.



(f) IDO-QFIRE Iris Matching ROC.

Figure 4.16: Iris recognition performance ROC's based on segmentation result evaluations. (a) Masek-WVU, (b) IDO-WVU (c) Masek-ICE, (d) IDO-ICE, (e) Masek-QFIRE, (f) IDO-QFIRE

solid curves (red) represent the matching performance for pairs of images that failed segmentation (ground truth). Finally, the dotted red curves show the matching performance for pairs of images that were evaluated as having failed segmentation boundaries. These graphs are useful in that we confirm two concepts. One, we confirm the previously accepted premise that matching performance is significantly impacted by segmentation results. Perhaps the more useful conclusion is that the proposed ensemble accurately evaluates overall segmentation results as the matching performance of the filtered results corresponding to the evaluated data closely resembles the matching performance of the filtered results for the ground truth data.

4.6 Chapter Summary

We presented an approach to automatically measure the results of iris segmentation algorithms. Scores are provided for the two boundaries relevant to the task: pupil segmentation and iris segmentation. We evaluated the approach using three algorithms across three publicly available data sets. The results indicate the approach is capable of arriving at segmentation scores suitable for evaluating both the success and failure of pupil or iris segmentation. Additionally, we present a ensemble based machine learning approach to arrive at an overall segmentation result which achieves an average classification accuracy of *92.52%* across all combinations of algorithms and data sets tested. We also presented one application of the proposed technique where the overall iris segmentation evaluation is used to filter iris recognition matching scores into correctly segmented and incorrectly segmented scoring bins. Here we confirmed that iris match scores hailing from images that were evaluated to have

good segmentation scores perform more accurately than pairs which were evaluated to have failed segmentation. While this is one application of the technique, the technique should prove useful in many other arenas such as iris quality metrics involving local analysis, image reacquisition, and a means to signal the need for more intensive segmentation processing for segmentation rectification.

Chapter 5

Segmentation Rectification

Recall that segmentation evaluation has many potential applications, not the least of which are related to image recapture or selectively filtering images that have been evaluated to be correctly segmented. The latter was demonstrated in Chapter 4. In this chapter we describe the more operationally relevant and certainly more attractive applications of segmentation evaluation. Notably, the scenarios involving the application of segmentation rectification.

5.1 Introduction

Segmentation rectification, simply put, is the process of re-segmenting the input iris image and serves as a potentially more attractive alternative to image recapture dependant upon the application. When segmentation fails, image recapture may not always be feasible for a given operational scenario. In particular, surveillance applications (e.g., iris at a distance) or portal scenarios which advocate high subject throughput cannot afford the opportunity for image recapture. As a result, segmentation rectification strategies must be explored. Here,

we enumerate in the following list several segmentation rectification strategies explored in this chapter.

1. ***Black box rectification*** $\sim S1$ - A simple yet potentially effective strategy may involve the application of a more computationally expensive segmentation algorithm, observed to provide better segmentation performance. However, such a strategy can be susceptible to making the same errors as the original and in time critical applications, just not as feasible. Furthermore, most commercial algorithms are proprietary, and obtaining additional, reportedly better, segmentation algorithms may be cost prohibitive, ultimately limiting the applicability or utility of said strategy. Nevertheless, this option is still certainly worth exploring. Here, the initial base segmentation algorithms are treated as a black box, providing only the segmentation boundaries as an initial segmentation output. When the segmentation evaluation ensemble detects a failed segmentation subsequent processing attempts to rectify said segmentation. Note that only the overall classification output from the ensemble is utilized.

2. ***Rectification through operator conditioning*** $\sim S2$ - The next strategy discussed in this chapter, S2, is fundamentally different from strategy S1. Here we assume that access to the internals of the initial base segmentation algorithm is available, specifically the search operator space. Thus, the segmentation evaluation measures can be applied straightforwardly to the search operators of the algorithm in question. That is, the evaluation measures may be utilized to condition the search operator such that a correct boundary is more likely to be chosen from the operator search space. This is

advantageous in that segmentation completes in only a single pass. As a result, this strategy is computationally more efficient than strategies S1 and S3.

3. ***Black box rectification with Segmentation Evaluation ~ S3*** - This strategy is a combination of S1 and S2, with the exception that operator conditioning is only applied when the ensemble flags a failed segmentation. Further, we utilize the individual C-SVM models (as well as their features) to indicate which boundary requires rectification. Since additional *a priori* information, related to the segmentation boundaries, is available through the individual evaluation measures, we can make a better judgement on how to rectify the incorrect segmentation. Given the proprietary nature of most vendor algorithms, it may not be possible to rectify segmentation boundaries with the same algorithm which provided the original segmentation. Like strategy S1, the initial base segmentation algorithms are treated as a black box, providing only the segmentation boundaries as an initial segmentation output. Thus, to rectify failed segmentation we build upon existing segmentation algorithms in the public domain, and modify them such that they take advantage of the information provided by the individual evaluation measures and the output of their respective models.

The aforementioned list of rectification strategies is by no means exhaustive but does serve as a starting point for exploring the potential of rectifying failed segmentation. Throughout the remainder of this chapter, we provide a look into the feasibility of segmentation rectification based on the aforementioned strategies. For the latter two strategies we make use of

the Masek and IDO segmentation algorithms combined with information generated by the segmentation evaluation block. We test the accuracy of each strategy using three different iris data sets. Additionally, we investigate the impact that each rectification strategy has on iris matching performance.

5.2 Segmentation Rectification Strategy - S1

The first rectification strategy, S1, that we explore is straightforward. Here, the segmentation evaluation block flags the utilization of a reportedly better but computationally expensive algorithm when the C-SVM ensemble detects an overall failed segmentation. We consider segmentation output from both the Masek and IDO algorithms as the initial base segmentation, treated as a black box. When a failed segmentation is detected via the ensemble, rectification is attempted with a secondary algorithm that is also considered a black box. For our experiments we chose the Zuo et al. [130] algorithm. Note that this strategy does not make use of the output from the individual evaluation measure models, only the overall output generated by the ensemble for evaluation purposes.

5.2.1 Experimental Results

Summary statistics are provided in Table 5.1 for each combination of data set and baseline algorithm. Note that the column, labeled “Evaluated-Failed”, represents the total number of detected failed segmentations from the ensemble, more than the reported number in the ground truth. This excess results from correct segmentation being falsely classified as failed segmentation, about 9.11% on average across all combinations of data sets and segmentation

(a)			(b)		
Masek - WVU			IDO - WVU		
	Evaluated-Failed	Failed Post-Rectification		Evaluated-Failed	Failed Post-Rectification
Frequency	976	32 (+14)	Frequency	468	23 (+48)
EER	32.07	6.49	EER	10.55	7.80
Mean Genuine Scores	.4026	.2861	Mean Genuine Scores	.3057	.2753
Mean Imposter Scores	.4569	.4519	Mean Imposter Scores	.4600	.4525

(c)			(d)		
Masek - ICE			IDO - ICE		
	Evaluated-Failed	Failed Post-Rectification		Evaluated-Failed	Failed Post-Rectification
Frequency	520	22 (+9)	Frequency	296	17 (+10)
EER	24.24	7.18	EER	16.07	11.13
Mean Genuine Scores	.3796	.2530	Mean Genuine Scores	.3368	.2638
Mean Imposter Scores	.4578	.4512	Mean Imposter Scores	.4593	.4490

(e)			(f)		
Masek - QFIRE			IDO - QFIRE		
	Evaluated-Failed	Failed Post-Rectification		Evaluated-Failed	Failed Post-Rectification
Frequency	447	100 (+1)	Frequency	207	77 (+8)
EER	46.60	28.34	EER	36.78	32.82
Mean Genuine Scores	.4210	.3630	Mean Genuine Scores	.3976	.3875
Mean Imposter Scores	.4359	0.4444	Mean Imposter Scores	.4457	.4449

Table 5.1: Summary statistics for rectification strategy S1. The “Evaluated-Failed” column represents the total number of detected failed segmentations via the ensemble, while the “Failed Post-Rectification” column corresponds to iris images whose boundaries could not be rectified. Clearly, rectification strategy S1 improves iris matching performance as the observed EER and mean genuine hamming distance scores for all data sets decreases post rectification.

algorithms. The “Failed Post-Rectification” column represents the number of segmentations that were not able to be rectified through the Zuo et al. algorithm. The number in parenthesis corresponds to the number of ground truth failed segmentations that the ensemble failed to detect outright, hence they did not have an opportunity to be rectified. Each table also reports the average hamming distance for genuine and imposter scores before and after segmentation rectification. In Figure 5.1, we provide ROC curves which characterize the performance of strategy S1 before and after segmentation rectification. Clearly, the improvement is more significant for the Masek algorithm than the IDO operator across all data sets. The IDO algorithm has fewer failed segmentation than the Masek algorithm. Furthermore, the characterization of failed segmentation produced by the IDO algorithm has more overlap

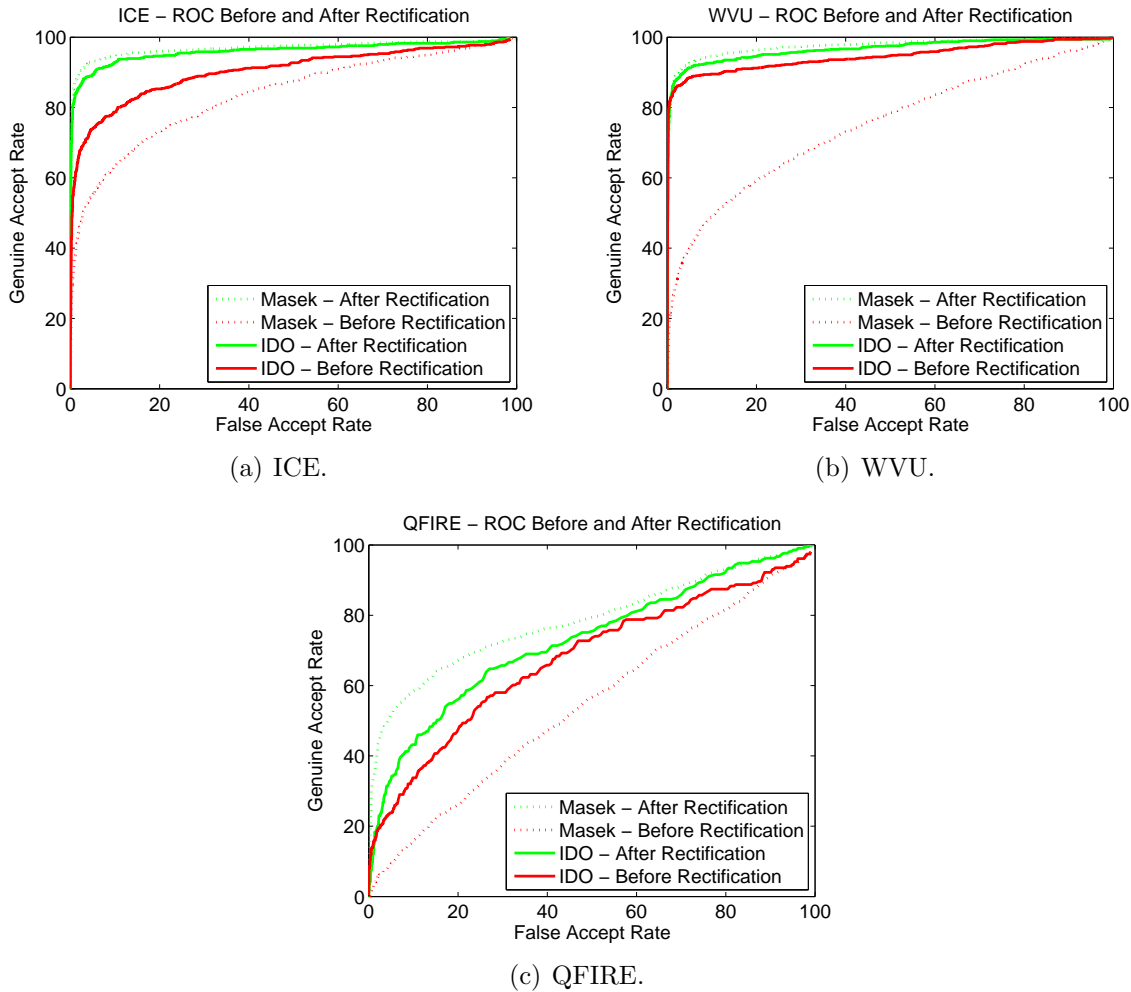


Figure 5.1: ROC curves characterizing the performance of strategy S1 rectification performance across all data sets. Rectification strategy S1 is observed to increase iris matching performance for all data sets. However, there is less observed improvement for the QFIRE data resulting from overlap in segmentation failures between the baseline and rectification segmentation algorithms.

with the characterization of correct segmentation across both data sets (see Figure 4.7 in Chapter 4) in comparison to the Masek algorithm. These tables and plots are informative in that we confirm the following concepts. First, we again illustrate that matching performance is significantly impacted by segmentation performance. This is more obvious with the Masek baseline algorithm since an overwhelming majority of segmentation failures produced by this algorithm can be characterized as grossly inaccurate (see 4.9(e) and 4.8(e) in Chapter 4). It is also important to note that while performance improved for both segmentation algorithms after rectification of the QFIRE data, it was not to the same extent as the other data sets, particularly for the IDO algorithm. The main reason for this is related to the overlap in failed segmentations across both algorithms. In other words, both the Zuo et al. and IDO algorithm had 77 images in common that failed segmentation (out of a total of 135 failures for the IDO algorithm), thus only 58 failures were rectified. In contrast, the Masek algorithm had 100 failures in common with the Zuo et al. algorithm but had more total failures than IDO (e.g., 423), hence 323 failures were successfully rectified.

5.3 Segmentation Rectification Strategy - S2

Rectification strategy S1 utilizes a computationally expensive segmentation algorithm to estimate new boundaries. While the results indicate the efficacy of this approach, having additional segmentation algorithms is not always feasible. Thus, a suitable alternative is to make a relatively weaker segmentation algorithm stronger. To that end, the next rectification strategy we experiment with conditions the operator space of the segmentation algorithm. More specifically, we wish test whether the additional information provided by the evaluation

measures can be utilized to improve segmentation performance. Note that the evaluation models are not utilized here. Instead, we generate the proposed evaluation measures and exploit them for operator conditioning. To test this strategy, we choose the IDO and Masek operators as the base segmentation algorithms. The Masek algorithm makes use of the Hough transform, H , as described in Chapter . The iris search process is modified such that the iris edge information is exploited to weight the maximum candidate for each possible search radius. Thus for the Masek algorithm the following modified Hough operator, \hat{H} , is utilized to estimate a new iris boundary. Mathematically, this is described as follows:

$$\max \hat{H}_{(x_c, y_c, r)} = \max [H_{(x_c, y_c, r)}] \times Lw_{(x_c, y_c, r)} \times Rw_{(x_c, y_c, r)}, \quad (5.1)$$

where

$$w_{(x_c, y_c, r)} = \frac{\sum_{i=1}^{|\theta|} \left[1 - \frac{\|\min \overline{D}_{\theta_i(x_c, y_c, r)} - I_{\theta_i(x_c, y_c, r)}\|}{\frac{N}{2}} \right]}{|\theta|}. \quad (5.2)$$

Here, \overline{D} represents the radial slice centered over the estimated iris boundary point, I , at angle θ_i . Note that $|\cdot|$ operator corresponds to cardinality while the $\|\cdot\|$ operator represents the Euclidean norm. Intuitively, weights are generated separately for the anatomical left, L , and right, R , hemisphere of the boundary, clock-wise from the estimated iris center. For the left hemisphere, $\theta = \{10^\circ, 15^\circ, 20^\circ\}$ while $\theta = \{170^\circ, 165^\circ, 160^\circ\}$ for the right hemisphere. Finally, the maximum of \hat{H} , yields an estimate for the new iris boundary conditioned via the proposed evaluation measures. Likewise, the search process of the pupil boundary is modified to include edge weights. The same angles are utilized as with the iris edge information.

Furthermore, we include the information generated by the pupil over/under-segmentation

$$\max \hat{H}_{(x_c, y_c, r)} = \max [H_{(x_c, y_c, r)}] \times Lw_{(x_c, y_c, r)} \times Rw_{(x_c, y_c, r)} \times P_{O_{(x_c, y_c, r)}} \times P_{U_{(x_c, y_c, r)}}, \quad (5.3)$$

scores, P_{Ow} and P_{Uw} , respectively. Note that these modifications are straightforward with the IDO algorithm. Rather than iterating through the Hough accumulator matrix, the maximum candidate boundary (iterating over each maximum r) generated by the IDO algorithm is weighted by the same evaluation measures in image space. The boundaries which maximize the IDO operator

$$\max_{(r, x_0, y_0)} \left| G_\sigma(r) * \frac{\partial}{\partial r} \oint_{r, x_0, y_0} \frac{I(x, y)}{2\pi r} ds \right| \times Lw_{(x_c, y_c, r)} \times Rw_{(x_c, y_c, r)}. \quad (5.4)$$

subject to the evaluation measures is selected as the boundary. Likewise, when searching for the pupil boundary, equation 5.4 is modified to include pupil over/under segmentation weights P_O and P_U .

5.3.1 Experimental Results

Summary statistics in terms of mean hamming distance scores and equal error rates are provided in Table 5.2 for each combination of data set and baseline algorithm. Note that after the data sets were segmented, the generated segmentations were ground truthed again according to the description provided in Chapter 4. These numbers are also provided in the table. The “Pre-Rectification” column corresponds to the baseline algorithms before their search operators were modified. The EERs before and after rectification for all

(a)			(b)		
Masek - WVU			IDO - WVU		
	Pre-Rectification	Post-Rectification		Pre-Rectification	Post-Rectification
Frequency	923	359	Frequency	408	128
EER	18.90	10.08	EER	6.14	4.00
Mean Genuine Scores	.3420	.3073	Mean Genuine Scores	.2882	.2784
Mean Imposter Scores	.4662	.4621	Mean Imposter Scores	.4678	.4618

(c)			(d)		
Masek - ICE			IDO - ICE		
	Pre-Rectification	Post-Rectification		Pre-Rectification	Post-Rectification
Frequency	379	100	Frequency	135	66
EER	8.20	2.62	EER	4.27	2.49
Mean Genuine Scores	.3095	.2909	Mean Genuine Scores	.2829	.2800
Mean Imposter Scores	.4628	.4632	Mean Imposter Scores	.4625	.4630

(e)			(f)		
Masek - QFIRE			IDO - QFIRE		
	Pre-Rectification	Post-Rectification		Pre-Rectification	Post-Rectification
Frequency	423	273	Frequency	136	119
EER	42.14	33.43	EER	23.68	22.89
Mean Genuine Scores	.4170	.4063	Mean Genuine Scores	.3795	.3789
Mean Imposter Scores	.4438	.4559	Mean Imposter Scores	.4572	.4579

Table 5.2: Summary statistics for rectification strategy S2. The “Pre-Rectification” column represents the total number of ground truth failed segmentations, while the “Post-Rectification” column corresponds to iris images whose boundaries could not be rectified through search operator conditioning.

algorithms clearly indicate the efficacy of introducing the evaluation metrics into the actual segmentation process. Looking at the WVU data tables, the EER decreases from *18.90* to *10.08* for the Masek algorithm while the EER for the IDO algorithm decreases from *6.14* to *4.00*. Note that improvement is mainly observed in the genuine scores while the imposter scores remain relatively stable. This was also the case for matching distributions generated by the first rectification strategy. Likewise, improvement is more pronounced for the Masek algorithm in comparison to IDO. This observation is not unexpected given that the Masek algorithm makes far more errors than the IDO algorithm and given the characterization of failed segmentation for the Masek algorithm with respect to the ground truth. This observation is further confirmed in Figure 5.2. Here we illustrate ROC performance before and after rectification for each combination of data set and segmentation algorithm.

5.4 Segmentation Rectification Strategy - S3

The last strategy we explore makes use of more information that the segmentation evaluation process affords. Here, we consider segmentation output from both the Masek and IDO algorithms as the initial base segmentation which is treated like a black box. When a failed segmentation is detected via the ensemble, we invoke a modified segmentation routine for rectification. Recall that the S1 rectification strategy utilizes a “better” segmentation algorithm for rectification but as a result is computationally more expensive than the initial base segmentations algorithms. As mentioned in the introduction, the availability of reportedly better algorithms is questionable. Similar to rectification S2, a potential alternative is to make a relatively weak segmentation algorithm stronger. That is, we modify the seg-

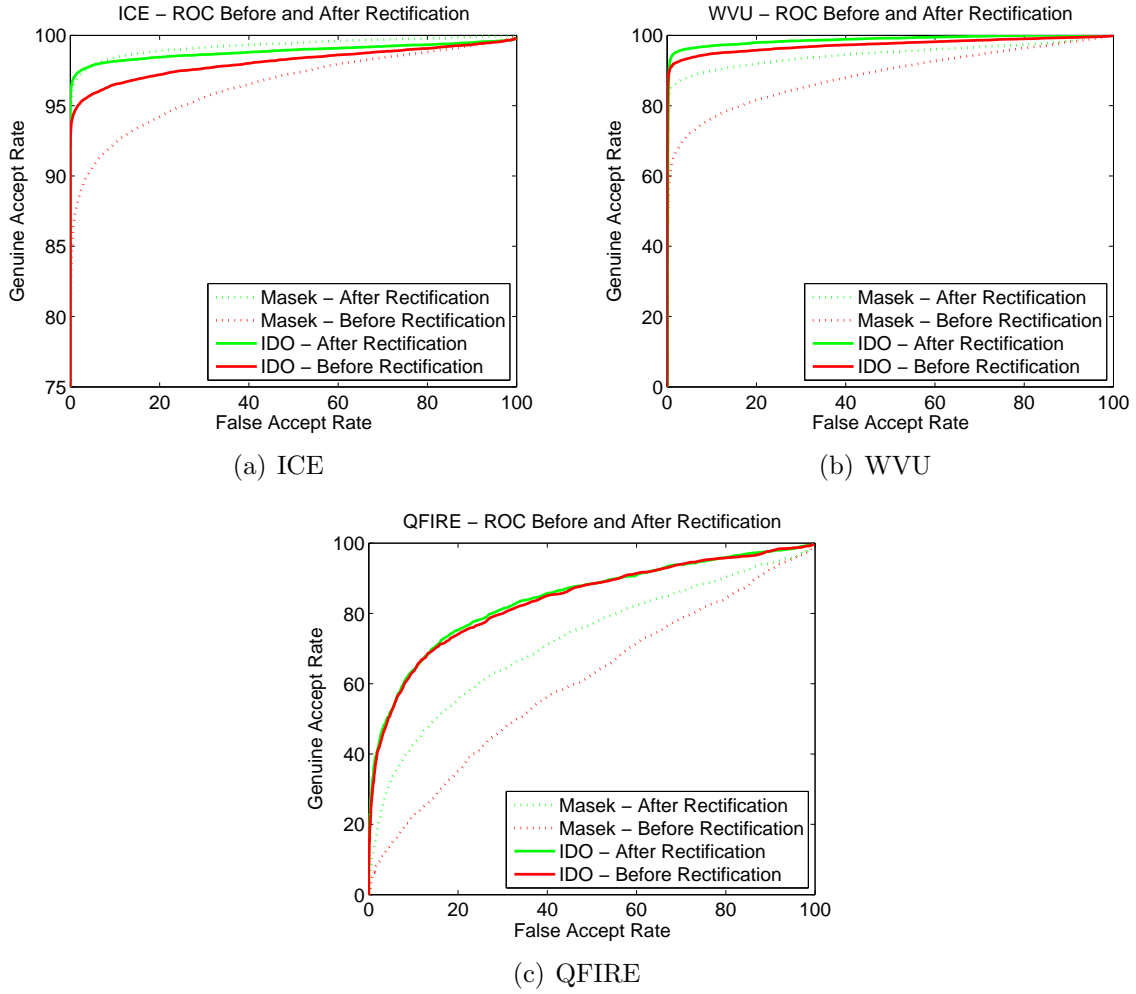


Figure 5.2: ROC curves characterizing the performance of strategy S2 rectification performance. Rectification strategy S2 is observed to increase iris matching performance for all data sets.

mentation routine to utilize information provided by the segmentation evaluation measures. Unlike the second strategy, we also include the individual C-SVM models generated from each feature to indicate which boundary requires rectification.

We adopt the Masek algorithm to serve as a skeleton for our modified segmentation routine. This approach utilizes a Hough transform based approach as described in Chapter 5.3. The first modification is related to the iris boundary search. That is, we use the C-SVM annular distance ratio and iris edge models to determine if the iris boundary is failed. If either one of these models indicate a failed boundary, then we utilize the modified Hough operator, \hat{H} , described in the previous section. After a new candidate iris boundary is estimated or if the models did not indicate a failed iris boundary, in which case the original estimated boundary is utilized, control proceeds to the pupil boundary. Here, the model generated from the pupil over/under segmentation features is evaluated. If the model does not detect an inferior boundary, processing is complete, otherwise we employ the following strategy for pupil rectification.

Recall that a likelihood ratio test is utilized to essentially label each pixel within (or surrounding for under-segmentation) the estimated pupil boundary as a true pupil pixel or non-pupil pixel. By doing so we can create an evaluation mask of the pupillary region, which can then be segmented to find a favorable pupil boundary. Intuitively, the mask is more homogeneous than the original image as it is only influenced by local image characteristics present within the segmentation boundaries being evaluated, thus providing a suitable candidate for edge map generation. Figure 5.3 provides an illustration of this process. In plot (a), we have an inferior segmentation. Edge map generation from the raw image, plot (d), yields noisy edge information. Note that the majority of edge information does not even

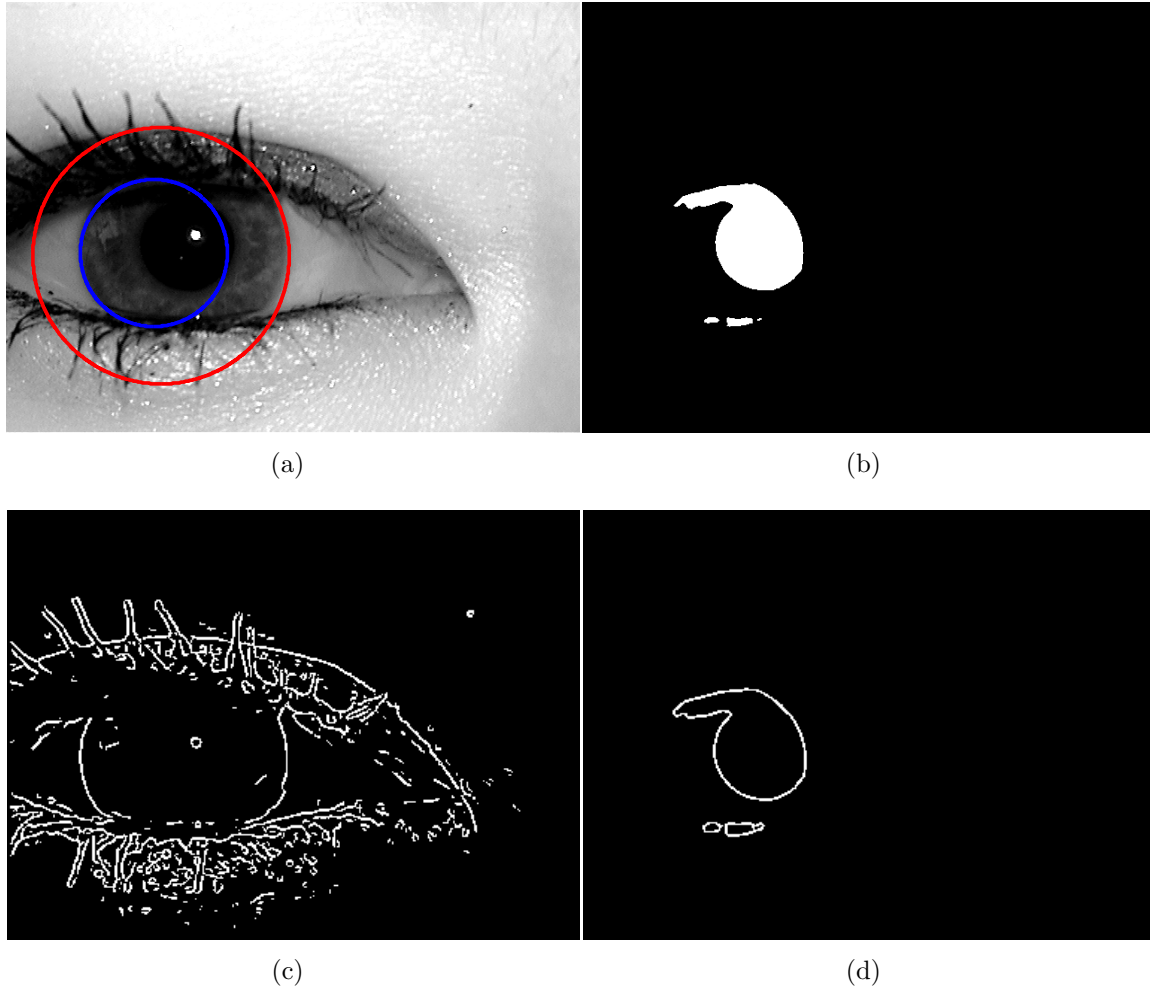


Figure 5.3: Illustration of pupil rectification process. (a) - Inferior segmentation overlayed on original image. (b) - Evaluation mask generated from the pupil over-segmentation and under-segmentation features. (c) - Edge map generated from original image. (d) - Edge map generated from the evaluation mask.

pertain to the pupil region. Plot (b) represents the evaluation mask generated by the likelihood ratio test statistic described in Chapter 4. Subsequent edge generation, illustrated in plot (d), yields a better candidate for locating the pupil boundary. Following this, a rectified

estimate of the pupil boundary is recovered via equations 2.8-2.10.

5.4.1 Experimental Results

The efficacy of the third rectification strategy is illustrated in Tables 5.3 and Figure 5.4. Here we provide summary statistics in terms of mean hamming distance scores and equal error rate. Also provided, are the ground truth results after the rectification process. As with the first two rectification strategies, performance of the Masek algorithm improves post rectification across all data sets tested. Improvement is observed mainly through a decreasing shift in genuine hamming distance distributions. However, the EER performance when experimenting on the WVU data is not as reasonable in comparison to the gain achieved through the first rectification strategy. That is, the EER decreases from *32.07* to *6.49* for the first rectification strategy in comparison to the third strategy. Unlike the Masek performance, the IDO results for the WVU data set do not indicate improvement. Specifically, the EER increases from *10.55* before rectification to *26.16* post rectification. After inspecting the segmentation boundaries post rectification, we observed *43* successful boundaries to actually flip. Recall that we are using the individual feature models to indicate whether a boundary needs rectified. Across all models, roughly *9.11%* of correct segmentation will be evaluated as failed. For the IDO model, this number is *108 (5.36%)*. Thus, *39%* of the evaluated correct segmentations flip as a result of the proposed rectification strategy. Furthermore, the characterization of the flipped boundaries poses a further problem. An overwhelming majority of failed segmentation for the Masek algorithm are observed to be grossly inaccurate in comparison to failures produced by the IDO algorithm. Since the S3 strategy utilizes the Hough based approach as a skeleton, IDO failures also take on this characterization. This

(a)			(b)		
Masek - WVU			IDO - WVU		
	Evaluated	Post-Rectification		Evaluated	Post-Rectification
Failed	976	403 (+14)	Failed	468	252 (+48)
EER	32.07	22.21	EER	10.55	26.16
Mean Genuine Scores	.4026	.3593	Mean Genuine Scores	.3057	.3569
Mean Imposter Scores	.4569	.4609	Mean Imposter Scores	.4600	.4601

(c)			(d)		
Masek - ICE			IDO - ICE		
	Evaluated	Post-Rectification		Evaluated	Post-Rectification
Failed	520	244 (+9)	Failed	296	212 (+10)
EER	24.24	7.54	EER	16.07	15.72
Mean Genuine Scores	.3796	.3305	Mean Genuine Scores	.3368	.3328
Mean Imposter Scores	.4578	.4634	Mean Imposter Scores	.4593	.4662

(e)			(f)		
Masek - QFIRE			IDO - QFIRE		
	Evaluated	Post-Rectification		Evaluated	Post-Rectification
Failed	447	256 (+1)	Failed	207	117 (+8)
EER	46.60	39.66	EER	36.78	37.32
Mean Genuine Scores	.4210	.4278	Mean Genuine Scores	.3976	.4094
Mean Imposter Scores	.4359	.4650	Mean Imposter Scores	.4457	.4522

Table 5.3: Summary statistics for rectification strategy S3.

observation also applies to the IDO performance for both the ICE and QFIRE data set but

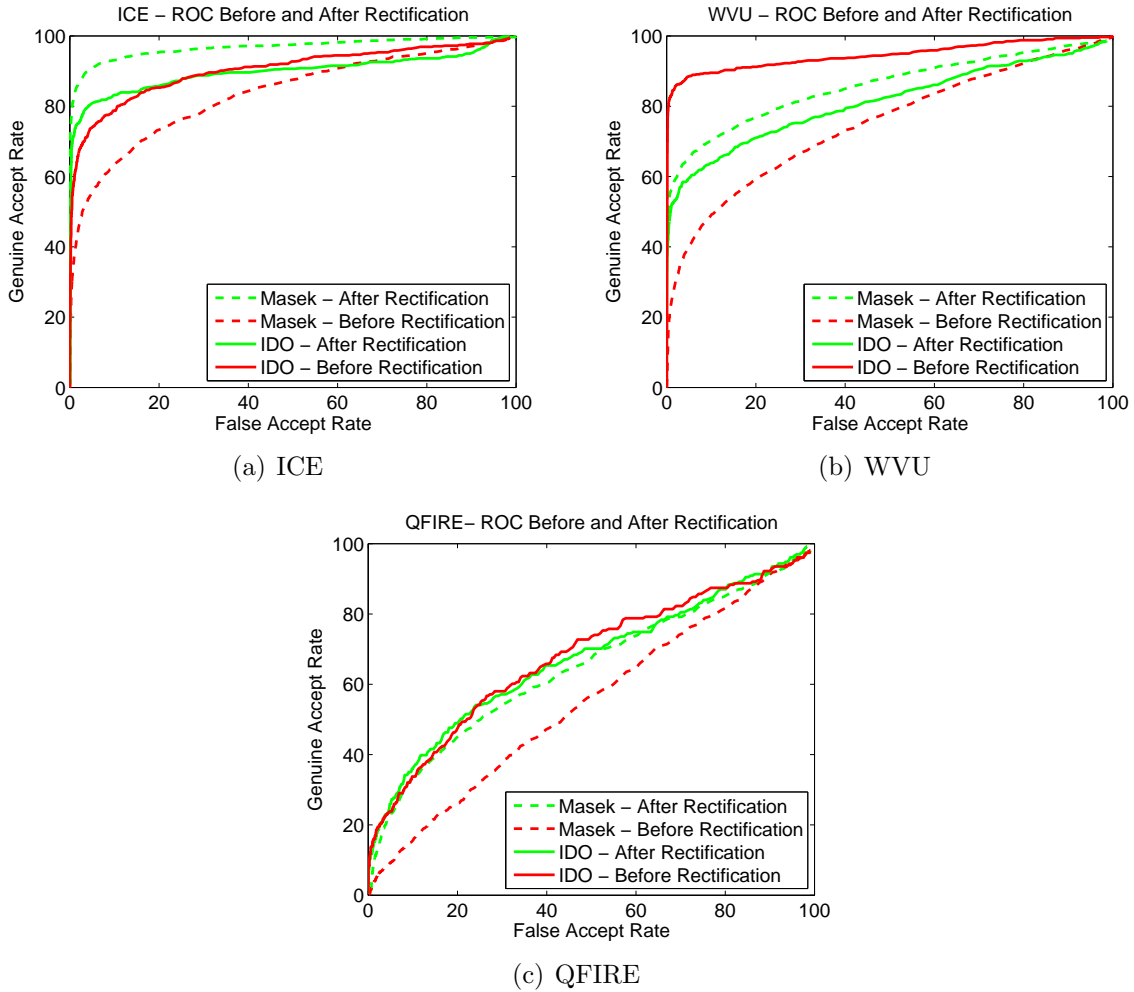


Figure 5.4: ROC curves characterizing the performance of strategy S3 rectification performance.

to a lesser extent. That is, only 22 and 18 of the evaluated correct segmentations flip out of a total of 171 (6.07%) and 79 (17.67%) for ICE and QFIRE respectively.

5.5 Chapter Summary

In this chapter we presented three strategies that automatically attempt to rectify inferior iris segmentation boundaries. The first technique utilizes the global ensemble output as an indicator to use a computationally more expensive segmentation algorithm when a failed segmentation is detected. While the results demonstrated the efficacy of the approach, obtaining such an algorithm for this task may be problematic given the proprietary nature of commercial algorithms. The second strategy explored in this chapter conditions the operator space of the segmentation algorithm. For the baseline segmentation algorithms tested, improvement was noted by a decreasing shift in the genuine hamming distance distributions. Besides illustrating the efficacy of the approach with respect to performance, the generality of the evaluation measures is also demonstrated. The last strategy we experiment with utilizes the evaluation features and their models to attempt rectification. Performance improvement was mainly observed with the baseline Masek segmentation in comparison to IDO segmentation as a result of correct segmentation boundaries flipping to failed.

Chapter 6

Iris Image Interoperability

6.1 Introduction

The main focus of this dissertation thus far has been related to the quality of the iris image, quality of the boundaries generated from segmentation, and on the rectification of segmentation evaluated to be inferior. In this last chapter, we look into the issue of iris image interoperability. We then illustrate how the application of hardware fingerprinting can be utilized to maintain iris match rates when images captured from distinct iris sensors are involved.

The field of iris recognition continues to proliferate, not only in algorithm development (to accommodate the non-ideal trend), but also in sensing technology. As new sensors are engineered, the question of interoperability arises. That is, are new sensors interoperable with existing sensors? This is a challenging problem for a number of reasons, not the least of which are related to the intended operational environment or even changes in the operational environment across different sessions of acquisition. Additionally, many factors may impact

the efficacy of iris matching performance. Such factors may include but are not limited to illumination (e.g., through dilation), contact lenses, and gaze (see Chapter 3). It is also reasonable to assume that the sensitivity of each sensor, with respect to the aforementioned factors, may vary according to the underlying design and optics of the sensor in question.

Several studies [22, 23, 105] have illustrated that iris matching performance is degraded when the matching process involves iris imagery acquired from different iris sensors. This degradation can potentially be mitigated or avoided entirely through the application of hardware fingerprinting. That is, we can avoid cross sensor matching by coupling the images to their respective sensors that they were acquired from. More specifically, hardware fingerprinting is the process of identifying the source hardware used to capture an image regardless of the image content. The primary method of identifying the source hardware from which an image originated is analyzing differences in images resulting from sensor imperfections [63]. However, the feasibility of this approach is limited by the availability of data. For instance, not having the same sensor imagery in both gallery/probe sets may confound match results. Further, even if coupling is possible, match rates may not be optimal if the imagery originates from differing acquisition environments. Nevertheless, hardware fingerprinting can still prove useful in those applications where sensor data is available in both gallery and probe sets.

In this chapter we investigate the notion of improving matching performance among 3 commercially available biometric iris sensors through the application of hardware fingerprinting. We first describe the proposed hardware fingerprinting technique and illustrate the efficacy on a chimerical data set comprised from imagery acquired from 7 different iris sensors. We then illustrate the problem of iris image interoperability by examining the cross

sensor matching performance between three commercial iris sensors. Finally, we show how the proposed hardware fingerprinting technique can be utilized to maintain iris matching performance.

6.2 Approach

Given a set of biometric sensors for which reference templates exist, we wish to identify an unknown sensor based on the resulting captured imagery. This problem can be formally stated as follows. We have a candidate set of sensors, S , where $S = \{s_1, s_2, \dots, s_n\}$ and n is the number of distinct sensors. For each sensor, $s_i \in S$, there is a set of training images, $P_i = \{p_1, p_2, \dots, p_k\}$ where k is the number of training images for sensor s_i . We wish to extract features, $\aleph = \Lambda(P_i)$, from the set of training images belonging to s_i such that it can be distinguished from all other sensors in the set S . Using features, \aleph , a reference template r_i , is generated for each s_i , resulting in a set of reference templates $R = \{r_1, r_2, \dots, r_n\}$. When an image is presented from an unknown sensor, we compute r_u from \aleph_u , in order to identify the unknown sensor, s_u . Next, a pairwise comparison of the unknown template r_u with all the reference templates $r_i \in R$ is performed. Sensor s_u is concluded as s_i for the r_i that is most similar to r_u .

The main point of concern is related to extracting features with the functional Λ . As discussed in the previous section, the majority of research in this domain utilizes features derived from imperfections in the sensor manufacturing process. In particular, the pixel non-uniformity (PNU) component of photo response non-uniformity noise results from pixels in the sensing array having unique sensitivities to light, and is arguably the most promising

approach to date. Therefore, we adopt the approach proposed by Lukas et al. in [63] as our choice for Λ .

6.2.1 Sensor Identification

The process of sensor identification consists of two integral steps: (1) Generating a noise reference pattern for each sensor; (2) Correlating noise residuals generated from test images to the aforementioned reference pattern(s).

1. ***Generate reference pattern.*** For each iris sensor, a reference pattern is calculated by taking an average of the noise residual estimates across multiple training images. This is mathematically characterized as follows:

$$\aleph_k = p^{(k)} - F(p^{(k)}), \quad (6.1)$$

$$\aleph_i = \frac{\sum_{k=1}^N \aleph_k}{N}. \quad (6.2)$$

Here, N represents the number of training images used to generate the reference pattern, \aleph_i (see figure 6.1 for examples). Noise residual, \aleph_k , is generated from training image $p^{(k)}$ while F represents a denoising filter. It should be noted that while F can represent any denoising filter, Lukas et al. found that a wavelet-based approach yielded the best results which is described later in this section.

2. ***Correlate test residuals to reference pattern(s).*** For each input test image, the noise residual, \aleph_k , is extracted and subsequently correlated with each reference pattern \aleph_i . Pearson's product-moment correlation coefficient is adopted which is described as

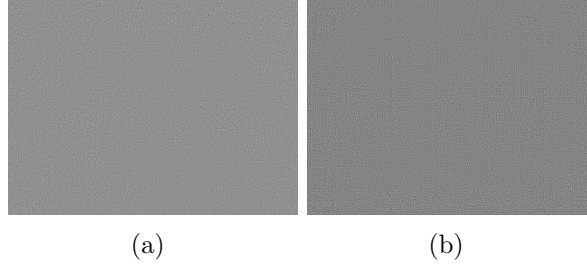


Figure 6.1: Illustration of reference templates generated from 5 training samples: (a) ICE-LG and (b) WVU-OKI.

follows:

$$\rho_i(\mathfrak{N}_{(k)}, \mathfrak{R}_i) = \frac{(\mathfrak{N}_{(k)} - \mathfrak{N}_{(k)}^-)(\mathfrak{R}_i - \mathfrak{R}_i^-)}{\|\mathfrak{N}_{(k)} - \mathfrak{N}_{(k)}^-\| \|\mathfrak{R}_i - \mathfrak{R}_i^-\|}. \quad (6.3)$$

6.2.2 Wavelet Based Denoising Algorithm

The wavelet based denoising algorithm, F , can be characterized into four steps:

1. **Wavelet Decomposition.** The original noisy image is decomposed into four levels utilizing wavelets, specifically, 8-tap Daubachies Quadratic Mirror Filters (QMF). The vertical, horizontal, and diagonal coefficients are denoted by $v(i, j)$, $h(i, j)$, and $d(i, j)$ respectively. Here (i, j) represents the coefficients for each pixel in each of the three sub-bands.
2. **Local MAP variance estimation.** In each sub-band, estimate the local variance of the noise-free image for each wavelet coefficient using MAP estimation for four sizes of a $W \times W$ neighborhood N , where $W \in \{3, 5, 7, 9\}$ as illustrated in the following equation.

$$\hat{\sigma}_W^2(i, j) = \max \left[0, \frac{1}{W^2} \sum_{(i, j) \in N} h^2(i, j) - \sigma_0^2 \right]. \quad (6.4)$$

Calculate the minimum of the four local variances,

$$\hat{\sigma}^2(i, j) = \min [\sigma_3^2(i, j), \sigma_5^2(i, j), \sigma_7^2(i, j), \sigma_9^2(i, j)] . \quad (6.5)$$

3. **Wiener Filtering.** The denoised wavelet coefficients are subsequently obtained after Wiener filtering.

$$h_{den}(i, j) = h(i, j) \frac{\hat{\sigma}^2(i, j)}{\hat{\sigma}^2(i, j) + \sigma_0^2} \quad (6.6)$$

4. **Repeat.** Steps 1-3 are repeated for each decomposition level and color channel. In [63], the authors used $\sigma_0^2 = 5$ in their experiments as do we in this work. Due to the imaging characteristics of near infrared iris sensors (typically monochrome CCD), the resulting imagery after capture is single channel or grayscale, therefore it is not necessary to perform Step 4 across multiple color channels.

6.3 Experimental Design

6.3.1 Chimerical Data Set Construction

To evaluate the efficacy of the proposed fingerprinting technique we first assembled a database composed of image subsets from seven publicly available iris databases. Specifically, we utilized subsets of imagery from ICE1.0, WVU Non-Ideal, WVU Off-Angle, CASIAv2-Device1, CASIAv2-Device2, CASIAv3-Lamp, and CASIAv3-Interval. Sample images and noise residuals can be found in table 6.2. The following provides a description of each database.

1. **ICE1.0** - This data set was collected and utilized for the Iris Challenge Evaluation

conducted by NIST [77]. Each image was collected with an LG IrisAccess EOU 2200 having a native resolution of 480x640. A total of 2953 images were collected from 244 classes. It is also important to note that this data was intentionally collected with a broad range of quality in mind.

2. **WVU Non-Ideal** - Collected with a hand held OKI Irispass-h device at a native resolution of 480x640 [90]. Like ICE1.0, this data set was intentionally collected with a range of quality (defocus, motion, non-uniform illumination, and occlusion). We utilized a subset of this database which was composed of 2424 images from 354 classes.
3. **WVU Off-Angle** - An EverFocus CCD camera with a native resolution of 480x640 was utilized for collection [90]. This data is composed of 146 iris classes with each class represented by 4 images captured at a yaw angle of 0° (2 samples), 15°, and 30° degrees.
4. **CASIAv2-Device1** - This subset is composed of 1200 images, collected from 60 classes. Images were captured with a hand held OKI Irispass-h device which has a native resolution of 480x640 [16].
5. **CASIAv2-Device2** - This subset is composed of 1200 images, collected from 60 classes. Images were captured with a proprietary device developed by CASIA having a native resolution of 480x640 [16].
6. **CASIAv3-Lamp** - Collected at the with a hand held OKI Irispass-h device [16] at a native resolution of 480x640. This data is composed of 16212 images, collected from 819 classes.

7. **CASIAv3-Interval** - Collected with a proprietary device developed by CASIA which has a native resolution of 280x320. Composed of 2639 images, collected from 395 classes [16].

From each iris database, we selected 200 images for our experiments. Selection was done such that a single image was chosen from each class (up to a max of 200 images). Multiple images were selected from each class for data sets where the total number of classes was less than 200. The success of the digital fingerprinting technique was tested while varying the amount of images used to generate reference patterns. A breakdown of the train and test scenarios is provided in Table 6.1. It is also important to note that we employed a

Train/Test Scenarios					
Train	4	8	16	32	64
Test	196	192	184	168	136

Table 6.1: Train and test scenarios for all sensors.

10-fold cross-validation framework for all seven sensors/data sets while testing the proposed technique. Therefore, the total number of tests for each sensor ranges from 1960 ($196 * 10$) to 1360 ($136 * 10$).

6.3.2 Interoperability - 2009 Biometric Collection Project Data Set

To illustrate the problem associated with iris image interoperability we rely on the iris image data collected within the 2009 Biometric Collection Project. Specifically, the iris data was collected from approximately 1000 subjects (resulting in 2000 classes) with three distinct















Dataset	Abbreviation	Sensor Model	Manufacturer	Image	Noise	Width	Height	Format
ICE1.0	ICE-LG	IrisAccess EOU 2200	LG			640	480	TIFF
WVU Non-Ideal	WVU-OKI	Irispass-h	OKI			640	480	BMP
WVU Off-Angle	WVU-EverFocus	Monochrome CCD	EverFocus			640	480	BMP
CASIAv2-Device1	CASIAv2-OKI	Irispass-h	OKI			640	480	BMP
CASIAv2-Device2	CASIAv2 _p	n/a	CASIA			640	480	BMP
CASIAv3-Lamp	CASIAv3-OKI	Irispass-h	OKI			640	480	JPEG
CASIAv3-Interval	CASIAv3 _p	n/a	CASIA			320	280	JPEG

Table 6.2: Illustration of sample images and corresponding noise residuals from each database.

commercially available iris sensors.

The first sensor used is the **LG-iCAM 4000**. This device is a dual iris capture sensor. It utilizes 12 NIR LED illuminators of varying wavelengths. For successful acquisition, stand-off distance is approximately 14 inches with the eyes centered over a reflective mirror. Five sessions were collected from each subject. Each session consists of 4 images captured at a resolution of 480×640 , the left and right iride, captured with the left and right illuminators. Note that in this study only images captured with corresponding illuminator are considered (e.g., Left eye with left illuminators and right eye with right illuminators). Further, variability was introduced after each session by requiring the subject to look away from the device prior to the next capture.

The second sensor used in this study is the **Crossmatch ISCAN2**. This device is a dual iris capture sensor, with similar usage to a pair of binoculars. Five sessions were collected from each subject, with each session consisting of a anatomical left and right iris acquired at a resolution of 480×480 .

The third sensor utilized is the **Aoptix Insight**. Like the other two, this device is a

dual iris capture sensor. For successful acquisition, the stand-off distance is approximately 2m with the subject looking at a small LCD on the device which provides a visual cue to “Look Here”. Five sessions were collected from each subject with each session consisting of the anatomical left and right eye. Note that this device has a capture volume of approximately $.75 \times 1m$ centered at the 2m mark. Thus, in order to increase variability across each session subjects moved to a different position within the capture volume. The first session was acquired with the subject centered in the rear of the capture volume. In contrast, the second session was acquired with the subject centered at the front of the capture volume. The third and fourth session consists of the subject standing on the right and left edge of the capture volume at 2m.

For training purposes we randomly selected 1 image from each class which resulted in a total of 2178 images. The success of the digital fingerprinting technique was tested while varying the amount of images used to generate reference patterns. A breakdown of the train and test scenarios is provided in Table 6.3.

Train/Test Scenarios				
Train	4	32	128	256
Test	2174	2146	2050	1922

Table 6.3: Train and test scenarios for the three commercial iris scanners.

It is also important to note that we employed a 10-fold cross-validation framework for the three sensors/data sets while evaluating the proposed technique. Therefore, the total number of tests for each sensor ranges from 21740 ($2174 * 10$) to 19220 ($1922 * 10$).

Classified Actual	ICE-LG	WVU-OKI	WVU-EverFocus	CASIAv3-OKI	CASIAv3 _p	CASIAv2-OKI	CASIAv2 _p
ICE-LG	1680	0	0	0	0	0	0
WVU-OKI	0	1680	0	0	0	0	0
WVU-EverFocus	9	0	1661	0	0	10	0
CASIAv3-OKI	0	0	0	1665	0	15	0
CASIAv3 _p	103	155	47	210	1009	82	74
CASIAv2-OKI	0	0	0	0	0	1680	0
CASIAv2 _p	0	0	0	0	0	0	1680

Table 6.4: Confusion matrix when training on 32 images per sensor.

6.4 Results - Hardware Fingerprinting

In this section we present results with respect to performing sensor identification at the unit level and combination of brand and unit levels. Recall that brand level sensor identification refers to identifying sensors manufactured by different vendors while unit level identification attempts to differentiate between sensors of the same model manufactured by the same vendor. In either case, we wish to distinguish which sensor the image was captured from. Therefore, a test noise residual is compared against reference patterns for each sensor from the data set in consideration. Results are illustrated in the form of match/non-match histograms, confusion matrices for specific train and test scenarios, and Cumulative Match Characteristic (CMC) curves.

6.4.1 Unit Level Sensor Identification

The first set of experiments were performed with WVU-OKI, CASIAv2-OKI, and CASIAv3-OKI in the context of unit level identification since they all were captured with an OKI sensor of the same brand and model. Figure 6.2 illustrates the difference in correlation between match and non-match comparisons of test noise residuals when considering the OKI sensor

from CASIAv3. Clearly, there is some overlap in both distributions. Specifically, 15 test residuals were misclassified as the CASIAv2-OKI device. Perfect separation was achieved when classifying residuals from the other two OKI devices, which is listed in table 6.4. In

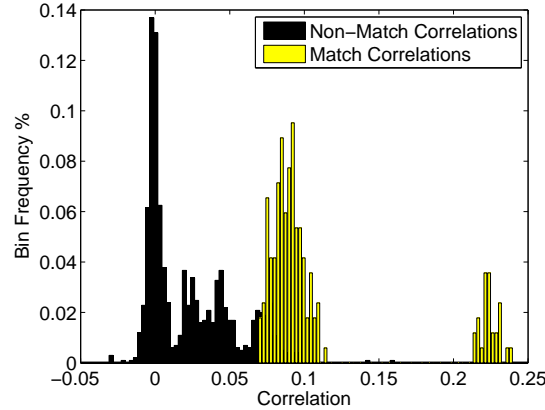


Figure 6.2: Example match and non-match distributions with 32 training images for CASIAv3-OKI.

Figure 6.3 we plot a CMC curve, which indicates overall accuracy, for all three OKI devices as a function of train/test sizes. When training on only 4 images per sensor, the rank one identification rate is around 86%. This is very encouraging considering the small amount of data utilized for training. This rate increases up to 99.75% when the train size increases to 64 images per sensor. Here the number of test residuals misclassified for the CASIAv3-OKI device reduces to seven.

6.4.2 Unit and Brand Level Sensor Identification

The following set of experiments explore performance when attempting sensor identification as a combination of the unit and brand levels for all sensors in this study. The ICE-LG,

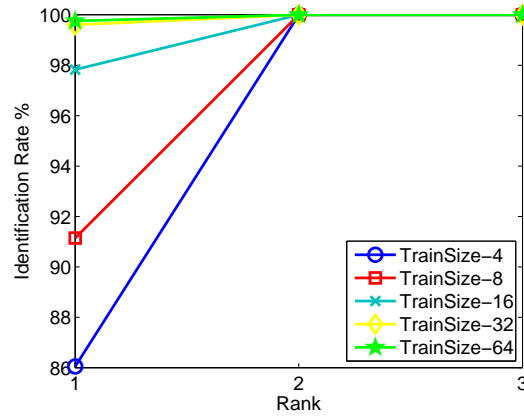


Figure 6.3: Unit Level Sensor identification as a function of training set size.

WVU-OKI, CASIAv2-OKI, and CASIAv2_p sensors provided the best performance as perfect separation was achieved (See table 6.4) when utilizing 32 training images per reference pattern. An example is provided in Figure 6.4 which illustrates the separation of match and non-match distributions for the ICE-LG sensor. On the other hand, perfect separation was not attainable for the WVU-EverFocus, CASIAv3-OKI, and CASIAv3_p test residuals. Specifically, few errors in classification were observed for the WVU-EverFocus and CASIAv3-OKI test residuals while CASIAv3_p proved the most challenging (see table 6.4). Here, misclassification of test residuals is spread almost uniformly across the remaining six sensors. This is also observed when utilizing the largest amount of train images to generate a reference pattern. There are a number of plausible explanations for this, not the least of which is JPEG compression (the images were released in JPEG format). Overall accuracy in terms of CMC performance across all sensors and train/test scenarios is characterized in Figure 6.5. Here, the rank one identification accuracy approaches 81% when generating a reference patterns from four training image samples. This number increases to 96% when

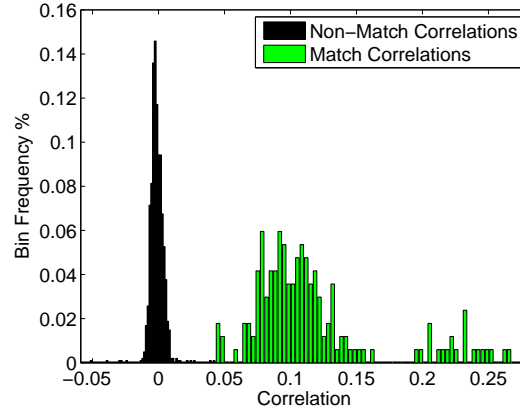


Figure 6.4: ICE1.0 example match and non-match distributions with 32 training images per sensor.

Classified Actual	ICE-LG	WVU-OKI	WVU-EverFocus	CASIAv3-OKI	CASIAv3 _p	CASIAv2-OKI	CASIAv2 _p
ICE-LG	1680	0	0	0	0	0	0
WVU-OKI	0	1387	0	293	0	0	0
WVU-EverFocus	15	9	1605	23	19	0	9
CASIAv3-OKI	0	0	0	1664	0	16	0
CASIAv3 _p	102	147	44	234	985	87	81
CASIAv2-OKI	0	0	0	1090	0	590	0
CASIAv2 _p	0	0	0	0	0	0	1680

Table 6.5: Confusion matrix when training on 32 images per sensor with JPEG compression (Quality=75).

considering 64 training image samples.

6.4.3 Impact of JPEG Compression

The next set of experiments analyze the impact JPEG compression has on the proposed sensor identification technique with respect to CMC performance. More specifically, this experiment analyzes performance when reference patterns are generated from iris images prior to compression while test residuals are generated from images compressed at JPEG

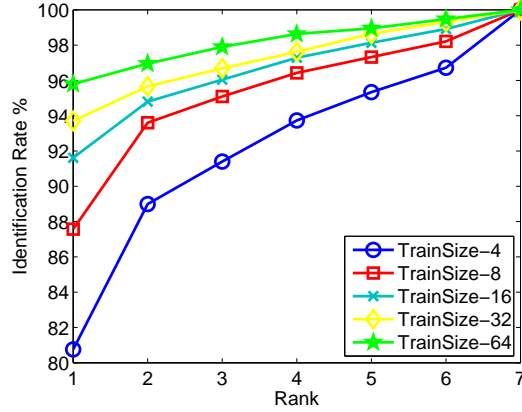


Figure 6.5: Cumulative match characteristic curve as a function of train/test scenarios when considering all seven sensors.

quality levels of 75, 50, and 35. The only exception is for CASIAv3-OKI and CASIAv3_p as the original image data was released in JPEG format. It is also important to note that the ISO/IEC 197946 Iris Image Data Standard [3] recommends a JPEG compression ratio of no more than 6 : 1. This ratio is approximately equal to JPEG quality levels between 90 – 95, but are outside the ranges tested within this paper. However, Daugman [31] has observed that iris recognition performance actually increases at JPEG quality 70. Therefore, the aforementioned JPEG quality levels (specifically 75), are reasonable in the current context.

Figure 6.6 illustrates the impact JPEG compression has on rank-1 and rank-2 identification performance when considering 32 train image samples to generate reference patterns. Most notably, accuracy degrades as the JPEG quality level decreases. This is further noted in the confusion matrix (for JPEG quality level 75) provided in table 6.5. Interesting, ICE-LG and CASIAv2_p test residuals remain unaffected by JPEG compression. On the other hand,

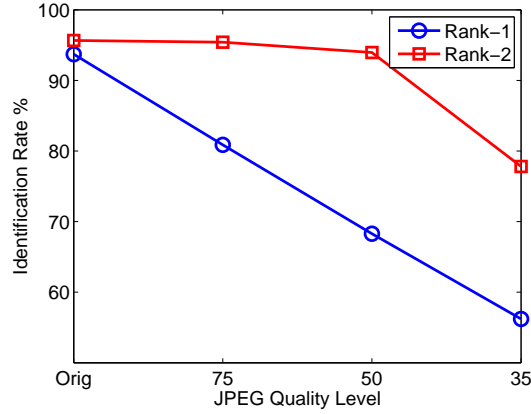


Figure 6.6: Rank-1 and Rank-2 Identification performance for all seven sensors as a function of JPEG quality level. In this illustration reference patterns were generated from 32 training images.

heavy degradation was noted during classification of the OKI test residuals, specifically, the WVU-OKI and CASIAv2-OKI residuals.

6.4.4 Interoperability - 2009 Biometric Collection Project Data Set

To illustrate the effectiveness of the hardware fingerprinting technique with respect to interoperability, we selected a subset of imagery from the 2009 Biometric Collection Project Data Set for match score generation. Notably, we selected 100 classes and their corresponding imagery from each sensor, resulting in a total of 500 images per sensor. Each image was segmented with the IDO algorithm while template generation was done with the Masek encoding scheme. The single sensor and cross-sensor matching results are illustrated in Figure 6.7(a). Here we observe that the ISCAN sensor provides the best matching performance in comparison to the LG and AOPTIX devices. This is not unexpected given the purported

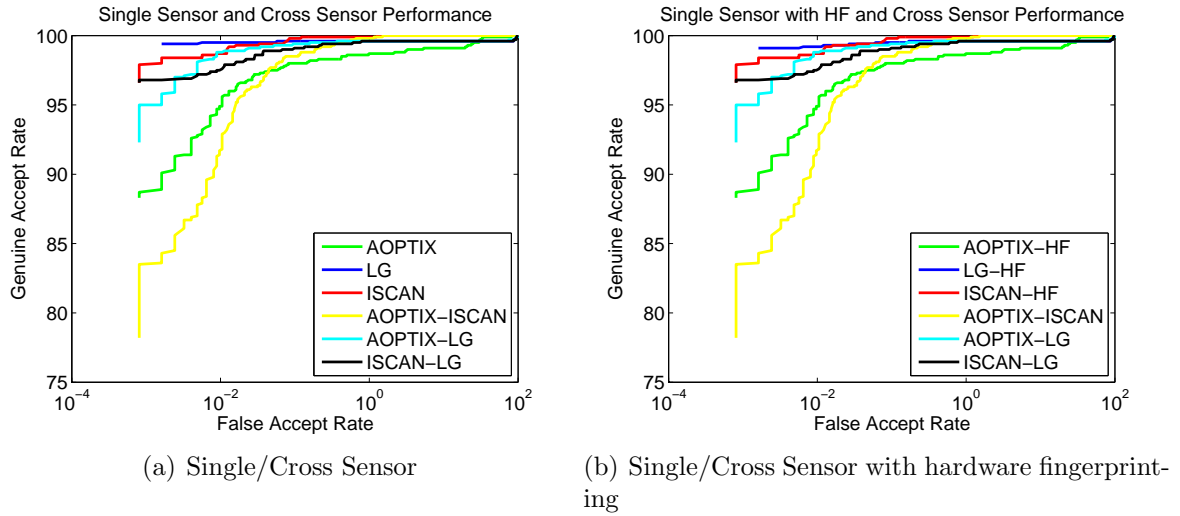


Figure 6.7: ROC curves characterizing the performance of single sensor and cross sensor performance. Note that in (b), hardware fingerprinting is used to ensure that matching is performed on imagery captured from the same sensor. As a result, single sensor performance is not degraded.

variability introduced during collection with the LG and AOPTIX devices. Also of interest is the cross-sensor ROC performance. With the exception of the AOPTIX imagery, cross-sensor performance is bounded above by its corresponding single sensor performance. The AOPTIX-LG combination provided the best cross sensor performance, followed by AOPTIX-ISCAN and ISCAN-LG. This is an unexpected result given that observations reported in [23]

EER performance			
	AOPTIX	LG	ISCAN
AOPTIX	1.3078 (1.3078)	0.3992	0.4811
LG	0.3992	0.3984 (0.3980)	0.5929
ISCAN	0.4811	0.5929	0.1114 (0.1118)

Table 6.6: EER performance for single sensor and cross sensor experiments. Note that the number in parenthesis corresponds to the ERR with hardware fingerprinting applied and probe data is unlabeled.

suggest the opposite. Here we observe that single sensor performance is not a predictor for cross sensor performance. To illustrate the impact hardware fingerprinting has on the matching experiments, a further experiment was conducted such that the gallery data from all sensors was available and the corresponding reference templates for each sensor. Next, the noise residuals were extracted from each unlabeled probe image and subsequently correlated to the reference templates. The probe was then matched to the gallery template which provided the highest correlation with respect to the reference template. The efficacy of this approach is demonstrated in Figure 6.7(b) and Table 6.6. More specifically, we should expect little change in the single sensor performance since the probe templates are being matched to their corresponding sensor gallery template.

6.5 Chapter Summary

In this chapter we investigated the issue of iris image interoperability for three commercial iris sensors. Our results indicate that cross sensor matching is bounded above by its corresponding single sensor performance. We also introduce the application of hardware fingerprinting to the matching process in order to couple probe imagery to their respective sensors that they were acquired from. Thus, the degradation observed from cross sensor matching can be mitigated. However, the feasibility of this approach is limited by the availability of data. For instance, not having the same sensor imagery in both enrollment/probe sets may confound match results, ultimately limiting the application of the proposed approach. Further, even if coupling is possible, match rates may be suboptimal if the imagery was acquired in different acquisition environments.

Chapter 7

Conclusions

7.1 Conclusions

In this dissertation we proposed a multi-faceted framework, for helping mitigate the aforementioned difficulties associated with non-ideal iris. Related to the acquisition process, we develop and investigate a comprehensive iris image quality metric that is predictive of iris matching performance. The metric is composed of photometric measures such as defocus, motion blur, and illumination, but is also composed of domain specific measures such as occlusion, and gaze angle. These measures are then combined through a fusion rule based on dempster-shafer theory. Related to iris segmentation, which is arguably one of the most important tasks in iris recognition, we developed several segmentation metrics which are used to evaluate the fidelity of the pupil and iris boundaries. Furthermore, we illustrated three methods which take advantage of the proposed segmentation metrics which can be utilized for rectifying incorrect segmentation boundaries. Finally, we look at the issue of iris image interoperability and demonstrate that techniques from the field of hardware fingerprinting

can be utilized to improve iris matching performance when images captured from multiple sensors are involved.

To reiterate, this dissertation provides the following contributions:

1. A comprehensive iris image quality metric that is predictive of iris matching performance. The metric is composed of photometric measures such as defocus, motion blur, and illumination, but is also composed of domain specific measures such as occlusion, and gaze angle. These measures are then combined through a fusion rule based on dempster-shafer theory.
2. Demonstration of the ability to reliably evaluate the success or failure of the iris segmentation routine. More specifically, the C-SVM model attains an average overall classification accuracy of 92.52% across all combinations of data sets and segmentation algorithms. The proposed approach analyzes the estimated segmentation boundaries for both pupil and iris, and derives geometric, shape, and edge metrics. These measures are then combined through a machine learning model which outputs a binary decision indicating the success or failure of estimated segmentation. These metrics are not tethered to any specific iris segmentation paradigm and as a result can be applied to many popular iris segmentation algorithms.
3. Demonstration of the ability to rectify incorrect segmentation boundaries when evaluated to be incorrect. We illustrate three methods which utilize the salient information generated by the segmentation evaluation methodology for the purpose of rectifying erroneous boundaries for both iris and pupil regions. The first method is a black box approach that utilizes the evaluation model to flag the use of a computationally more

expensive segmentation routine to rectify inferior segmentation. The second method utilizes the evaluation measures by conditioning the search operator of the baseline segmentation algorithms. Finally, the third method is a pseudo-combination of the first two methods in that we utilize the segmentation evaluation measures to condition and rectify inferior boundaries by augmenting a relatively weaker segmentation algorithm in order to make it stronger. The first strategy provided the best performance with respect to average percent decrease in EER, 42.82%, across all combinations of data sets and algorithms. However, this strategy relies on a “ringer” segmentation algorithm which may not be operationally feasible. In contrast, the second strategy, provided a modest average percent decrease in EER, 35.88%. Furthermore, it does not rely on any additional segmentation algorithms. Instead, the evaluation features are utilized which are not tied to any specific segmentation paradigm.

4. Demonstration of the application of hardware fingerprinting to iris sensors for increasing sensor interoperability. When images captured from multiple sensors are involved, we show that the application of hardware fingerprinting can be utilized to improve or at the very least, stabilize matching performance.

7.2 Future Work

While the underpinnings of this work provide several contributions that assist in mitigating the difficulties associated with non-ideal iris recognition, there are additional aspects of this work that will require additional investigation moving forward.

With respect to image quality, there is need for additional domain specific quality mea-

asures that correlate well with iris matching performance. In support of this notion, the NIST IQCE [100] evaluation(s) provide a wealth of information related to this effort. Specifically, adoption of morphological characteristics, which measure the shape of the pupillary and iris boundaries, may prove beneficial when attempting to estimate the quality of an iris image. Further, it is likely that image quality will need to be tightly coupled with segmentation rectification. This is motivated by the notion that even if segmentation boundaries are correctly rectified, the presence of noise within the iris region may impede the ability to successfully match.

Related to segmentation evaluation, the analysis presented in Chapter 4 indicates that characterization of segmentation failures is dependent upon the chosen segmentation algorithm. For instance, the Hough based algorithms utilized in this work had a tendency to generate grossly inaccurate boundaries upon failure as opposed to the IDO algorithm. This is particularly important because the proposed features which are utilized to characterize segmentation failure may not be informational with the advent of new state-of-art segmentation algorithms. While a number of recent works in the iris recognition literature exploit and improve upon the Hough transform or IDO operator, it would be beneficial to evaluate the ability of the proposed features to generalize beyond either foundational technique, specifically active contours [29, 98]. Another avenue of potential work, is related to extending the proposed evaluation framework to eyelid detection. While non-trivial, the edge based measures proposed in Chapter 4 could be extended and modified to evaluate arcuate boundaries typically generated for upper and lower eyelid occlusion.

Related to the topic of segmentation rectification, search operator conditioning and the invocation of a more comprehensive segmentation algorithm (when segmentation was eval-

uated to be incorrect) were observed to provide the best overall performance. However, the latter rectification strategy has limited utility, as this requires access to multiple segmentation algorithms which must be at least partially independent with respect to their failures. Further, the former strategy requires access to the internals of the segmentation algorithm while the majority of commercial iris algorithms are black boxes. The last rectification strategy employed provided inconclusive results. In particular, performance improved when rectifying failed segmentations from the Hough based algorithm. However this result did not extend to the IDO algorithm. We observed that one of the primary reasons for this is related to correct segmentation falsely classified as incorrect, flipping. That is, it was correct prior to rectification (e.g., verified by ground truth). Aside from building a stronger ensemble model with lower classification error rates, multi-pass rectification strategies could be explored.

While research is advancing in the field of iris image interoperability, additional effort is required to increase or maintain matching performance with images captured from different iris sensors. In particular, the proposed approach is limited by the availability of sensor specific data in both probe and gallery repositories. This requirement is not feasible in operational environments. A potential avenue for improvement may attempt to only match images that have been “binned” by similar image characteristics (e.g. matching imagery that have similar levels of noise, blur, occlusion, etc).

Bibliography

- [1] CASIA iris image database (ver 1.0). <http://www.sinobiometric.com/casiairis.html>.
- [2] Wvu off-angle iris biometric dataset collection. www.citer.wvu.edu, 2003.
- [3] Information technology, biometric data interchange format: Iris image data. Standard 19794-6, ISO/IEC, 2005.
- [4] A. Abhyankar and S. Schuckers. Active shape models for effective iris segmentation. *SPIE Biometric Technology for Human Identification III*, 6202, April 2006.
- [5] F. Bashir, P. Casaverde, D. Usher, and M. Friedman. Eagle-eyes: A system for iris recognition at a distance. In *IEEE Conference on Technologies for Homeland Security*, pages 426–431, May 2008.
- [6] F. Bashir, D. Usher, P. Casaverde, and M. Friedman. Video surveillance for biometrics: Long-range multi-biometric system. In *IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance*, pages 175 –182, September 2008.
- [7] C. Belcher and Y. Du. A selective feature information approach for iris image-quality measure. *IEEE Transactions on Information Forensics and Security*, 3(3):572 –577, September 2008.
- [8] R. M. Bolle, S. Pankanti, and N. K. Ratha. Evaluation techniques for biometrics-based authentication systems (firr). *Proc. 15th International Conference on Pattern Recognition (ICPR)*, 2:831 – 837, Sept 2000.

- [9] B. Bonney, R. Ives, D. Etter, and Y. Du. Iris pattern extraction using bit planes and standard deviations. In *Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 582–586, November 2004.
- [10] K. W. Bowyer, K. Hollingsworth, and P. J. Flynn. Image understanding for iris biometrics: A survey. *Computer Vision and Image Understanding*, 110(2):281–307, 2008.
- [11] C. Boyce. Multispectral iris recognition analysis: Techniques and evaluation. Master’s thesis, West Virginia University, Fall 2006.
- [12] C. Boyce, A. Ross, M. Monaco, L. Hornak, and X. Li. Multispectral iris analysis: A preliminary study. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW ’06. Conference on*, june 2006.
- [13] R. Broussard, L. Kennell, D. Soldan, and R. Ives. Using artificial neural networks and feature saliency techniques for improved iris segmentation. In *International Joint Conference on Neural Networks*, pages 1283–1288, August 2007.
- [14] M. Burge and M. Monaco. Multispectral iris fusion for enhancement, interoperability, and cross wavelength matching. In *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XV*, volume 7334, pages 733410 – 733418. SPIE, 2009.
- [15] T. Camus and R. Wildes. Reliable and fast eye finding in close-up images. In *Proceedings. 16th International Conference on Pattern Recognition*, volume 1, pages 389 – 394, 2002.
- [16] CASIA. Casia iris image database. Online, <http://biometrics.idealtest.org/>.
- [17] C. Chang and C. J. Lin. Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 27(2):1–27, 2011.
- [18] Y. Chen, S. Dass, and A. Jain. Localized iris quality using 2-D wavelets. *International Conference on Biometrics*, pages 373–381, 2006. Hong Kong, China.
- [19] Y. Chen, S. C. Dass, and A. K. Jain. Localized iris image quality using 2-d wavelets. In *International Conference on Biometrics*, pages 373–381, 2006.

- [20] C.-T. Chou, S.-W. Shih, W.-S. Chen, V. Cheng, and D.-Y. Chen. Non-orthogonal view iris recognition system. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(3):417–430, March 2010.
- [21] I. E. Comission. Safety of laser products part1: Equipment classification and requirements. Technical Report 3, IEC, 2007.
- [22] R. Connaughton, A. Sgroi, K. Bowyer, and P. Flynn. A cross-sensor evaluation of three commercial iris cameras for iris biometrics. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*, pages 90–97, june 2011.
- [23] R. Connaughton, A. Sgroi, K. Bowyer, and P. Flynn. A multialgorithm analysis of three iris biometric sensors. *Information Forensics and Security, IEEE Transactions on*, 7(3):919–931, june 2012.
- [24] H. S. M. Coxeter. *Projective Geometry*. Springer Verlag, 2nd edition, 2003.
- [25] S. Crihalmeanu, A. Ross, R. Govindarajan, L. Hornak, and S. Schuckers. A centralized web-enabled multimodal biometric database. *Biometric Consortium Conference (BCC)*, Sept 2004.
- [26] J. Daugman. Statistical richness of visual phase information: Update on recognizing persons by iris patterns. *International Journal of Computer Vision*, 45:25–38, October 2001.
- [27] J. Daugman. How iris recognition works. In *Proceedings International Conference on Image Processing*, volume 1, pages I–33 – I–36 vol.1, 2002.
- [28] J. Daugman. How iris recognition works. *IEEE Trans. on Circuits and Systems for Video Technology*, 14(1):21–30, January 2004.
- [29] J. Daugman. New methods in iris recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 37(5):1167–1175, October 2007.

- [30] J. Daugman and C. Downing. Epigenetic randomness, complexity and singularity of human iris patterns. *Proceedings of the Royal Society (London): B. Biological Sciences*, 268:1737–1740, 2001.
- [31] J. Daugman and C. Downing. Effect of severe image compression on iris recognition performance. *IEEE Transactions on Information Forensics and Security*, 3(1):52–61, March 2008.
- [32] J. G. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15:1148–1161, November 1993.
- [33] J. G. Daugman. Biometric personal identification system based on iris analysis. United States Patent 5,291,560, March 1994.
- [34] V. Dorairaj, N. Schmid, and G. Fahmy. Performance evaluation of non-ideal iris based recognition system implementing global ica encoding. In *IEEE International Conference on Image Processing*, volume 3, September 2005.
- [35] C. Fancourt, L. Bogoni, K. Hanna, Y. Guo, R. Wildes, N. Takahashi, and U. Jain. Iris recognition at a distance. In T. Kanade, A. Jain, and N. Ratha, editors, *Audio- and Video-Based Biometric Person Authentication*, volume 3546 of *Lecture Notes in Computer Science*, pages 1–13. Springer Berlin / Heidelberg, 2005.
- [36] X. Feng, C. Fang, X. Ding, and Y. Wu. Iris localization with dual coarse-to-fine strategy. In *18th International Conference on Pattern Recognition*, volume 4, pages 553–556, 2006.
- [37] L. Flom and A. Safir. Iris recognition system. United States Patent 4,641,349, February 1987.
- [38] I. O. for Standardization and I. E. Commission. Biometric data interchange formats - part 6: Iris image data. 19794-6, ISO/IEC, March 2004. Draft Version.
- [39] P. Grother, E. Tabassi, G. W. Quinn, and W. Salamon. Irex 1: Performance of iris recognition algorithms on standard images. Presentation, NIST, 2009.

- [40] G. Guo, M. Jones, and P. Beardsley. A system for automatic iris capturing. Technical Report TR2005-044, Mitsubishi Electric Research Laboratories, 2005.
- [41] L. Guo. *Software Quality and Reliability Prediction Using Dempster-Shafer Theory*. PhD thesis, West Virginia University, Morgantown, 2004.
- [42] X. Guo, G. Yang, and Y. Yin. Sensor interoperability of fingerprint segmentation: An empirical study. In *Information Engineering and Computer Science, 2009. ICIECS 2009. International Conference on*, pages 1–4, dec. 2009.
- [43] Z. He, T. Tan, and Z. Sun. Iris localization via pulling and pushing. In *Proceedings of the 18th International Conference on Pattern Recognition, ICPR 2006*, pages 366–369, Washington, DC, USA, 2006. IEEE Computer Society.
- [44] K. Hollingsworth, K. Bowyer, and P. Flynn. The best bits in an iris code. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(6):964–973, June 2009.
- [45] K. Hollingsworth, K. W. Bowyer, and P. J. Flynn. Pupil dilation degrades iris biometric performance. *Comput. Vis. Image Underst.*, 113(1):150–157, 2009.
- [46] K. Hollingsworth, K. W. Bowyer, and P. J. Flynn. Pupil dilation degrades iris biometric performance. *Computer Vision and Image Understanding*, 113(1):150–157, 2009.
- [47] Y.-P. Huang, S.-W. Luo, and E.-Y. Chen. An efficient iris recognition system. In *International Conference on Machine Learning and Cybernetics*, volume 1, pages 450 – 454, 2002.
- [48] A. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):4–20, January 2004.
- [49] N. Kalka, N. Bartlow, and B. Cukic. An automated method for predicting iris segmentation failures. In *IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–8, September 2009.

- [50] N. Kalka, J. Zuo, N. Schmid, and B. Cukic. Estimating and fusing quality factors for iris biometric images. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 40(3):509–524, May 2010.
- [51] N. D. Kalka, J. Zuo, N. A. Schmid, and B. Cukic. Image quality assessment for an iris biometric. *Proc. of the 2006 SPIE conference on Biometric Technology for Human Identification III*, 6202:62020D–1 – 62020D–11, April 2006.
- [52] B. Kang and K. Park. A study on iris image restoration. *Proc. Audio and Video-Based Biometric Person Authentication*, 3546:31–40, July 2005.
- [53] B. Kang and K. Park. A study on iris image restoration. *Proc. Audio and Video-Based Biometric Person Authentication*, 3546:31–40, July 2005.
- [54] R. Kerekes, B. Narayanaswamy, J. Thornton, M. Savvides, and B. V. K. V. Kumar. Graphical model approach to iris matching under deformation and occlusion. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 0:1–6, 2007.
- [55] E. Krichen, S. Garcia-Salicetti, and B. Dorizzi. A new probabilistic iris quality measure for comprehensive noise detection. In *First IEEE International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–6, September 2007.
- [56] R. D. Labati and F. Scotti. Noisy iris segmentation with boundary regularization and reflections removal. *Image and Vision Computing*, 28(2):270–277, 2010. Segmentation of Visible Wavelength Iris Images Captured At-a-distance and On-the-move.
- [57] S. Lee. Quality of iris segmentation as a predictor of verification performance. Master’s thesis, West Virginia University, August 2007.
- [58] P. Li and X. Liu. An incremental method for accurate iris segmentation. In *International Conference on Pattern Recognition*, pages 1–4, December 2008.
- [59] P. Lili and X. Mei. The algorithm of iris image preprocessing. In *Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, pages 134–138, October 2005.

- [60] X. Liu, K. W. Bowyer, and P. J. Flynn. Experiments with an improved iris segmentation algorithm. *IEEE Workshop on Automatic Identification Advanced Technologies*, 0:118–123, 2005.
- [61] X. Liu, K. W. Bowyer, and P. J. Flynn. Iris recognition and verification experiments with improved segmentation method. *Proc. Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, October 2005.
- [62] Y. Liu, S. Yuan, X. Zhu, and Q. Cui. A practical iris acquisition system and a fast edges locating algorithm in iris recognition. In *Proceedings of the 20th IEEE Conference on Instrumentation and Measurement Technology*, volume 1, pages 166 – 168, May 2003.
- [63] J. Lukas, J. Fridrich, and M. Goljan. Digital camera identification from sensor pattern noise. *IEEE TIFS*, 1(2):205–214, June 2006.
- [64] L. Ma, T. Tan, Y. Wang, and D. Zhang. Personal identification based on iris texture analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1519 – 1533, December 2003.
- [65] L. Masek. Recognition of human iris patterns for biometric identification. Master’s thesis, The University of Western Australia, 2003.
- [66] L. Masek. Recognition of human iris patterns for biometric identification. B.S. Thesis, The University of Western Australia, <http://www.csse.uwa.edu.au/pk/studentprojects/libor/>, 2003.
- [67] J. Matey, O. Naroditsky, K. Hanna, R. Kolczynski, D. LoIacono, S. Mangru, M. Tinker, T. Zappia, and W. Zhao. Iris on the move: Acquisition of images for iris recognition in less constrained environments. *Proceedings of the IEEE*, 94(11):1936–1947, November 2006.
- [68] J. R. Matey, R. Broussard, and L. Kennell. Iris image segmentation and sub-optimal images. *Image and Vision Computing*, 28(2):215 – 222, 2010.

-
- [69] J. R. Matey, O. Naroditsky, K. Hanna, R. Kolczynski, D. Loiacono, S. Mangru, M. Tinker, T. Zappia, and W. Zhao. Iris on the move: Acquisition of images for iris recognition in less constrained environments. *Proc. of the IEEE*, 94:1936 – 1947, November 2006.
- [70] D. E. Maurer and J. P. Baker. Fusing multimodal biometrics with quality estimates via a bayesian belief network. *Pattern Recogn.*, 41(3):821–832, 2008.
- [71] M. Mladenovski. Information fusion schemes for real time risk assessment in adaptive control systems. Master’s thesis, West Virginia University, 2004.
- [72] S. Modi, S. Elliott, and H. Kim. Statistical analysis of fingerprint sensor interoperability performance. In *Biometrics: Theory, Applications, and Systems, 2009. BTAS ’09. IEEE 3rd International Conference on*, pages 1 –6, sept. 2009.
- [73] R. Murphy. Adaptive rule of combinations for observations over time. *Fusion and Integration for Intelligent Systems*, pages 125 – 131, December 1996.
- [74] R. Murphy. Dempster-shafer theory for sensor fusion in autonomous mobile robots. *IEEE Trans. On Robotics and Automation*, 14(2):197–206, April 1998.
- [75] K. Nandakumar, Y. Chen, S. Dass, and A. Jain. Likelihood ratio-based biometric score fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):342–347, Feb. 2008.
- [76] E. Newton and P. Phillips. Meta-analysis of third-party evaluations of iris recognition. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 39(1):4–11, January 2009.
- [77] NIST. Iris challenge evaluation. Online, <http://iris.nist.gov/ice/> 2005.
- [78] P. Phillips, W. Scruggs, A. OToole, P. Flynn, K. Bowyer, C. Schott, and M. Sharpe. Ice 2006 report: Frvt 2006 and ice 2006 large-scale results. *NISTIR 7408*, 2007.

- [79] N. Poh, J. Kittler, and T. Bourlai. Improving biometric device interoperability by likelihood ratio-based quality dependent score normalization. In *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, pages 1–5, sept. 2007.
- [80] H. Proenca and L. Alexandre. Ubiris: A noisy iris image database. *Proc. of the International Conference on Image Analysis and Processing*, pages 970 – 977, 2005.
- [81] H. Proenca. Iris recognition: On the segmentation of degraded images acquired in the visible wavelength. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8):1502–1516, August 2010.
- [82] H. Proenca and L. Alexandre. Iris segmentation methodology for non-cooperative recognition. *IEE Proceedings - Vision, Image, and Signal Processing*, 153(2):199 – 205, April 2006.
- [83] H. Proenca and L. A. Alexandre. A method for the identification of noisy regions in normalized iris images. In *Proceedings of the 18th International Conference on Pattern Recognition, ICPR '06*, pages 405–408, Washington, DC, USA, 2006. IEEE Computer Society.
- [84] H. Proenca and L. A. Alexandre. Toward noncooperative iris recognition: A classification approach using multiple signatures. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(4):607–612, 2007.
- [85] H. Proenca and L. A. Alexandre. Toward noncooperative iris recognition: A classification approach using multiple signatures. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(4):607–612, 2007.
- [86] A. Ross. Iris recognition: The path forward. *IEEE Computer*, 43(2):30–35, February 2010.
- [87] A. Ross and A. Jain. Biometric sensor interoperability: A case study in fingerprints. In *in Fingerprints, Appeared in Proc. of International ECCV Workshop on Biometric Authentication*, pages 134–145, 2004.

-
- [88] A. Ross and R. Nadgir. A thin-plate spline calibration model for fingerprint sensor interoperability. *Knowledge and Data Engineering, IEEE Transactions on*, 20(8):1097–1110, aug. 2008.
- [89] A. Ross, R. Pasula, and L. Hornak. Exploring multispectral iris recognition beyond 900nm. In *Biometrics: Theory, Applications, and Systems, 2009. BTAS '09. IEEE 3rd International Conference on*, pages 1–8, sept. 2009.
- [90] S. S. S. Crihalmeanu, A. Ross and L. Hornak. A protocol for multibiometric data acquisition, storage and dissemination technical report. *West Virginia University, Lane Department of Computer Science and Electrical Engineering*, 2007.
- [91] S. I. Sarnoff. Iom glance compact system. <http://www.sarnoff.com/products/iris-on-the-move/compact-system>, 2008.
- [92] N. Schmid and F. Nicolo. A method for selecting and ranking quality metrics for optimization of biometric recognition systems. *Computer Vision and Pattern Recognition Workshop*, 0:126–133, 2009.
- [93] S. Schuckers, P. M. Lopez, P. Johnson, N. Sazonova, F. Hua, R. Lazarick, C. Miles, E. Talbassi, E. Sazonov, A. Ross, and L. Hornak. Quality - face / iris research ensemble (q-fire) dataset overview. Technical report, Clarkson University, Dept of Electrical and Computer Engineering, 2010.
- [94] S. Schuckers, N. Schmid, A. Abhyankar, V. Dorairaj, C. Boyce, and L. Hornak. On techniques for angle compensation in nonideal iris recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 37(5):1176–1190, October 2007.
- [95] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, 1982.
- [96] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
- [97] S. Shah and A. Ross. Iris segmentation using geodesic active contours. *Trans. Info. For. Sec.*, 4:824–836, December 2009.

- [98] S. Shah and A. Ross. Iris segmentation using geodesic active contours. *IEEE Transactions on Information, Forensics, and Security*, 4:824–836, December 2009.
- [99] H. Sung, J. Lim, J. hyun Park, and Y. Lee. Iris recognition using collarette boundary localization. In *Proceedings of the 17th International Conference on Pattern Recognition*, volume 4, pages 857 – 860 Vol.4, August 2004.
- [100] E. Tabassi, P. Grother, and W. Salamon. Iqce: Performance of iris image quality assessment algorithms. NIST Interagency Report 7820, 2011.
- [101] T. Tan, Z. He, and Z. Sun. Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition. *Image and Vision Computing*, 28(2):223 – 230, 2010.
- [102] A. Technologies. Insight sd 2-meter iris recognition system. <http://www.aoptix.com/iris-recognition/product/insight-sd>, 2008.
- [103] Q.-C. Tian, Q. Pan, Y.-M. Cheng, and Q.-X. Gao. Fast algorithm and application of hough transform in iris segmentation. In *Proceedings of 2004 International Conference on Machine Learning and Cybernetics*, volume 7, pages 3977 – 3980, August 2004.
- [104] R. Unnikrishnan, C. Pantofaru, and M. Hebert. Toward objective evaluation of image segmentation algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29:929–944, June 2007.
- [105] V. S. Valencia. Iris recognition interoperability and image quality. Presentation, Authenti-Corp, July 2007.
- [106] M. Vatsa, R. Singh, and A. Noore. Unification of evidence-theoretic fusion algorithms: A case study in level-2 and level-3 fingerprint features. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 39(1):47–56, January 2009.
- [107] M. Vatsa, R. Singh, A. Noore, and M. M. Houck. Quality-augmented fusion of level-2 and level-3 fingerprint information using dsm theory. *Int. J. Approx. Reasoning*, 50(1):51–61, 2009.

-
- [108] J. Wan, X. He, and P. Shi. An iris image quality assessment method based on laplacian of gaussian operation. In *IAPR Conference on Machine Vision Applications*, 2007.
- [109] Z. Wei, T. Tan, and Z. Sun. Nonlinear iris deformation correction based on gaussian model. In S.-W. Lee and S. Li, editors, *Advances in Biometrics*, volume 4642 of *Lecture Notes in Computer Science*, pages 780–789. Springer Berlin Heidelberg, 2007.
- [110] Z. Wei, T. Tan, Z. Sun, and J. Cui. Robust and fast assessment of iris image quality. *International Conference on Biometrics*, pages 464–471, 2006.
- [111] F. Wheeler, A. Perera, G. Abramovich, B. Yu, and P. Tu. Stand-off iris recognition system. In *IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 1–7, October 2008.
- [112] R. Wildes. Iris recognition: an emerging biometric technology. *Proceedings of the IEEE*, 85(9):1348–1363, September 1997.
- [113] R. Wildes. Iris recognition: An emerging biometric technology. *Proceedings of the IEEE*, 85(9):1348–1363, September 1997.
- [114] R. Wildes, J. Asmuth, G. Green, S. Hsu, R. Kolczynski, J. Matey, and S. McBride. A system for automated iris recognition. In *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*, pages 121–128, December 1994.
- [115] R. P. Wildes, J. C. Asmuth, G. L. Green, S. C. Hsu, R. J. Kolczynski, J. R. Matey, and S. E. McBride. A machine-vision system for iris recognition. *Machine Vision and Applications*, 9:1–8, 1996.
- [116] H. J. Wyatt. A minimum-wear-and-tear meshwork for the iris. *Vision Research*, 40(16):2167–2176, 2000.
- [117] G. Xu, Z. Zhang, and Y. Ma. Automatic iris segmentation based on local areas. In *International Conference on Pattern Recognition*, volume 4, pages 505–508, 2006.

- [118] X. Yuan and P. Shi. A non-linear normalization model for iris recognition. In S. Z. Li, Z. Sun, T. Tan, S. Pankanti, G. Chollet, and D. Zhang, editors, *Advances in Biometric Person Authentication*, volume 3781 of *Lecture Notes in Computer Science*, pages 135–141. Springer Berlin Heidelberg, 2005.
- [119] A. Zaim, M. Quweider, J. Scargle, J. Iglesias, and R. Tang. A robust and accurate segmentation of iris images using optimal partitioning. In *Proceedings of the 18th International Conference on Pattern Recognition*, ICPR '06, pages 578–581, Washington, DC, USA, 2006. IEEE Computer Society.
- [120] G. Zhang and M. Salganicoff. Method of measuring the focus of close-up image of eyes. Technical Report 5953440, United States Patent, 1999.
- [121] H. Zhang, J. E. Fritts, and S. A. Goldman. Image segmentation evaluation: A survey of unsupervised methods. *Computer Vision and Image Understanding*, 110(2):260 – 280, 2008.
- [122] Y. Zhang, X. F. Lu, H. L. Lu, and W. Liu. Iris image quality assessment based on fpga coprocessor. In *International Communication Conference on Wireless Mobile and Computing*, pages 555 –558, December 2009.
- [123] Z. Zhou, Y. Du, and C. Belcher. Transforming traditional iris recognition systems to work on non-ideal situations. In *IEEE Workshop on Computational Intelligence in Biometrics: Theory, Algorithms, and Applications*, pages 1–8, April 2009.
- [124] X. Zhu, Y. Liu, X. Ming, and Q. Cui. A quality evaluation method of iris images sequence based on wavelet coefficients in 'region of interest'. *Proc. of the 4th Int'l Conf. on Computer and Information Technology*, pages 24–27, September 2004.
- [125] J. Zuo, N. Kalka, and N. Schmid. An iris segmentation procedure for unconstrained subject presentation. *Proc. of the Biometric Consortium Conference*, 2006.

- [126] J. Zuo, F. Nicolo, and N. Schmid. Cross spectral iris matching based on predictive image mapping. In *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, pages 1 –5, sept. 2010.
- [127] J. Zuo and N. Schmid. An automatic algorithm for evaluating the precision of iris segmentation. In *2nd IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 1 –6, October 2008.
- [128] J. Zuo and N. Schmid. An automatic algorithm for evaluating the precision of iris segmentation. *International Conference on Biometrics: Theory, Applications and Systems*, pages 1–6, 29 2008-Oct. 1 2008.
- [129] J. Zuo and N. Schmid. Global and local quality measures for nir iris video. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 120 –125, 20-25 2009.
- [130] J. Zuo and N. A. Schmid. On a methodology for robust segmentation of nonideal iris images. *Trans. Sys. Man Cyber. Part B*, 40:703–718, June 2010.